


A pink piggy bank is the central focus, positioned on a large pile of gold coins. A single gold coin is captured in mid-air, just above the slot on the piggy bank's back. The scene is brightly lit, creating a warm, yellowish glow. The piggy bank has a classic design with a snout, ears, and a small tail.

**Prêt à dépenser**

Projet 7 - Raphaël Koudache--Louvet

An aerial, top-down view of a suburban neighborhood. The houses are mostly two-story with red-tiled roofs, some with solar panels. There are many green trees and several blue swimming pools. A paved road runs diagonally from the bottom left towards the top right. The overall scene is bright and clear.

**Analyse de la solvabilité des demandeurs de crédits immobiliers**  
**Classification binaire**  
**Réalisation et déploiement d'un dashboard à destination du banquier**



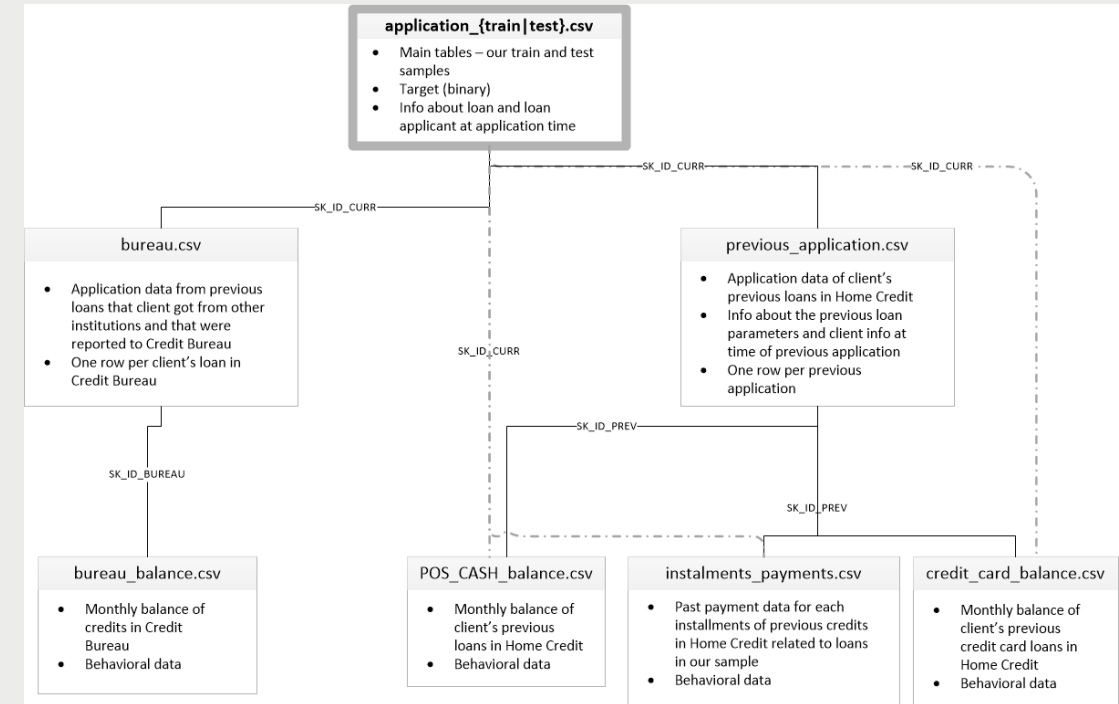
A large, heavy-duty metal vault door, likely made of steel, is shown in a partially open position. The door features a complex internal mechanism with multiple locking bolts and a central locking assembly. The door is set into a concrete wall. The text "I - Analyse Exploratoire des données (EDA)" is overlaid on the image.

# I - Analyse Exploratoire des données (EDA)

# Vue d'ensemble

Données fragmentées  
(plusieurs .csv)

Kernel Kaggle (Imputations,  
Aggrégations, Jointure)

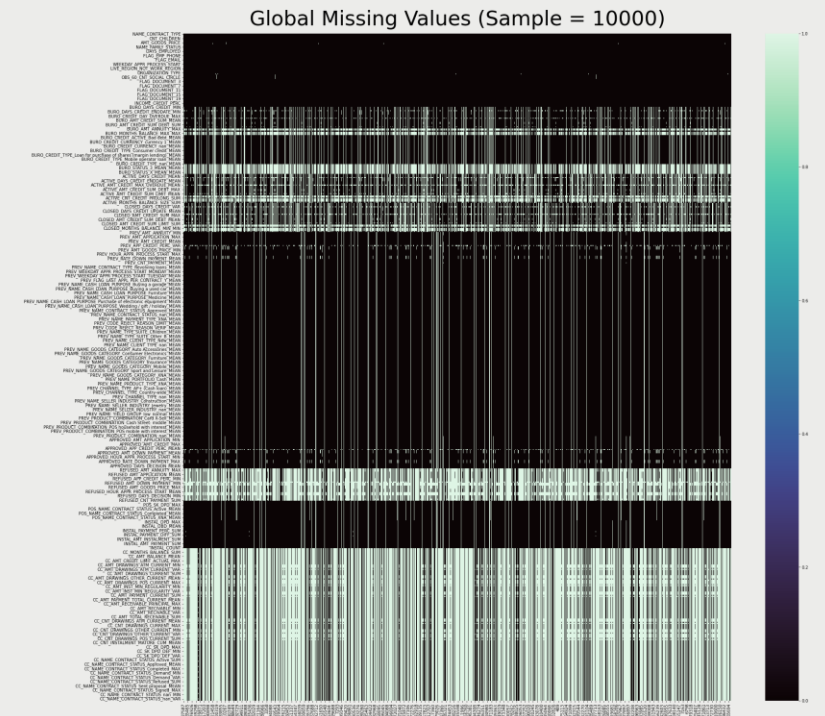
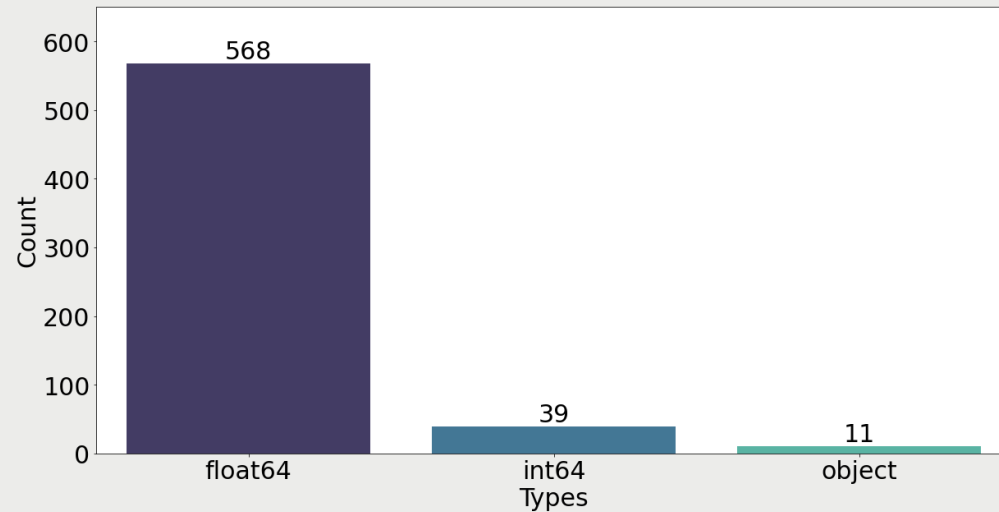




# Vue d'ensemble

Shape (307511,620)

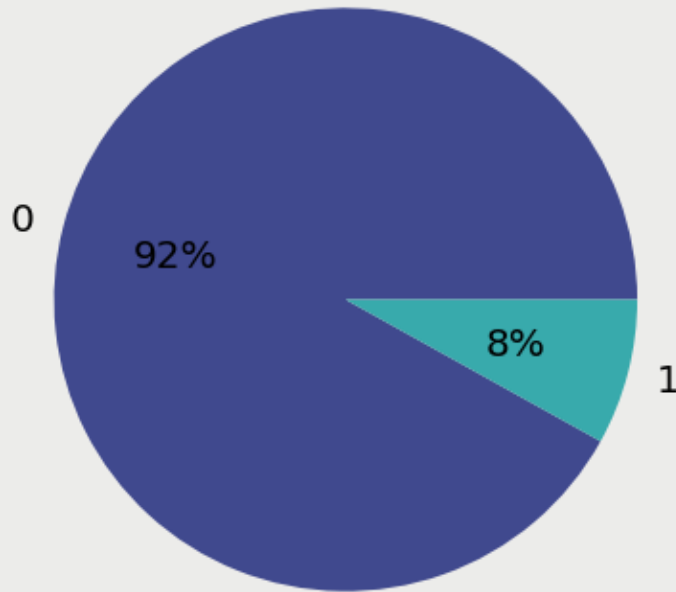
Types of features count



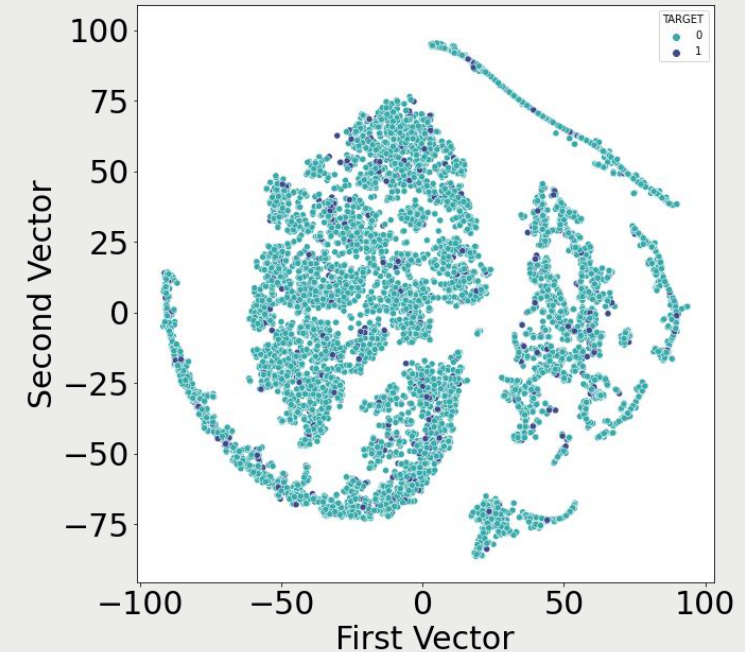


## Des cibles éparses et non balancées

# Target Distribution



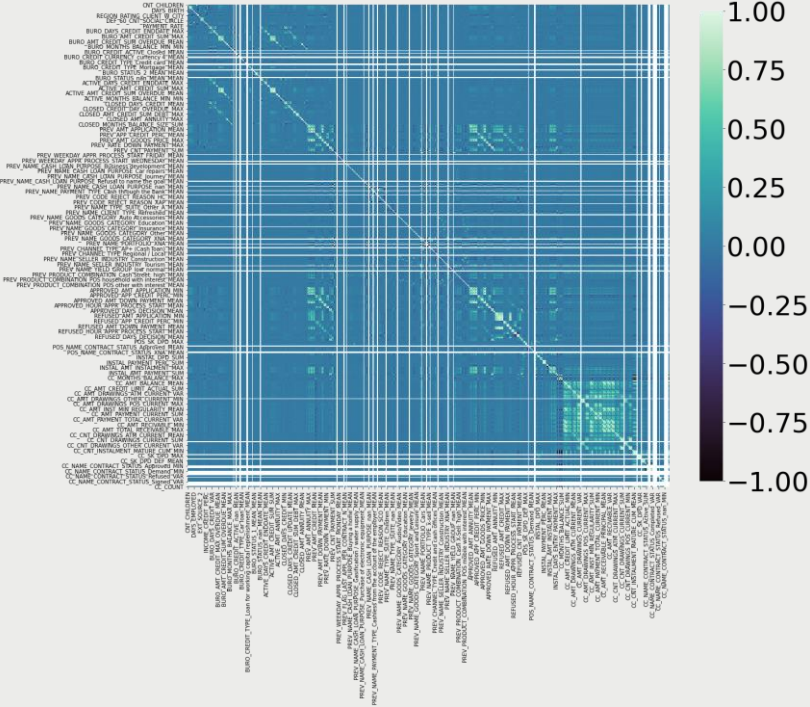
T-SNE 2 components (sample = 10000)



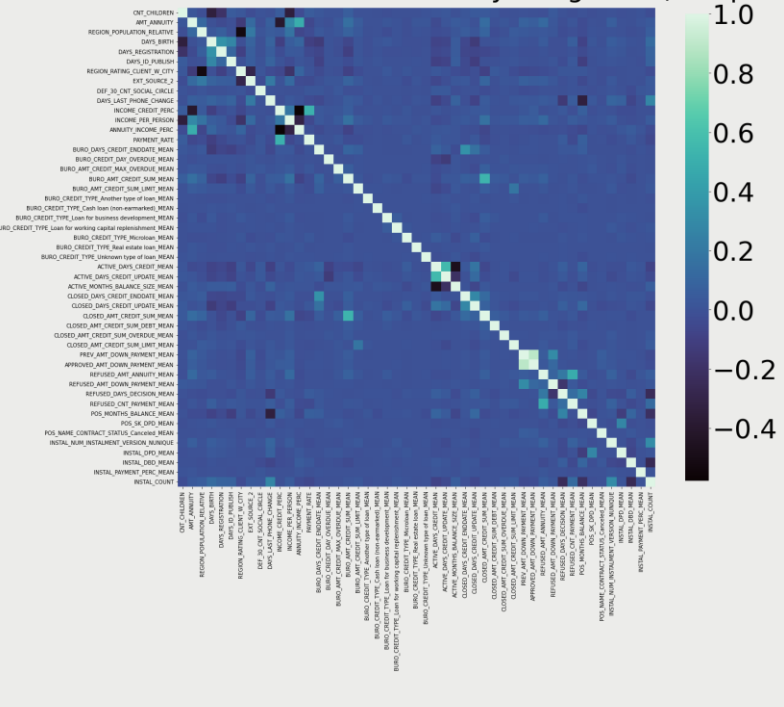


# Une Multicolinéarité forte à traiter

Pearson's coefficients (Sample = 10000)



Pearson's coefficients after multicollinearity mitigation (Sample = 10000)



Variance Inflation Factor, Tests Chi deux, ANOVA, Kruskal-Wallis



## II - Modélisation

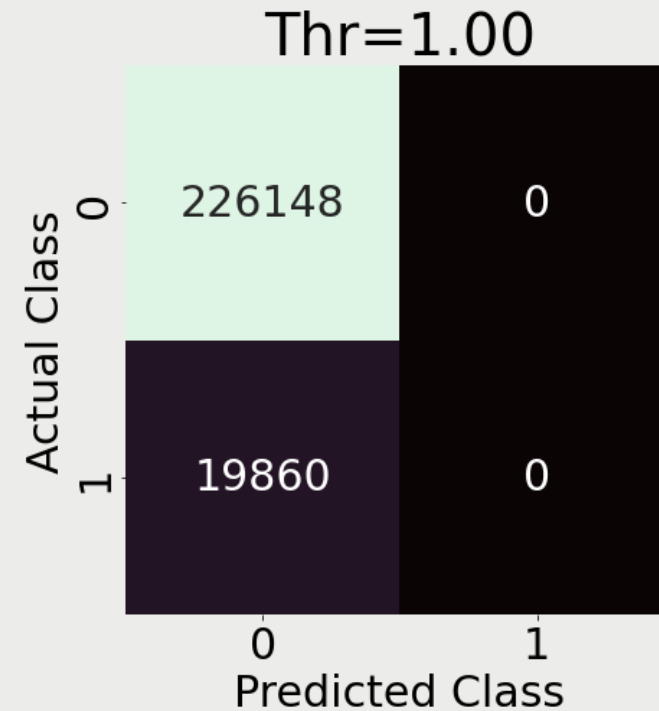




## Choix d'une métrique

Le paradoxe de l'accuracy

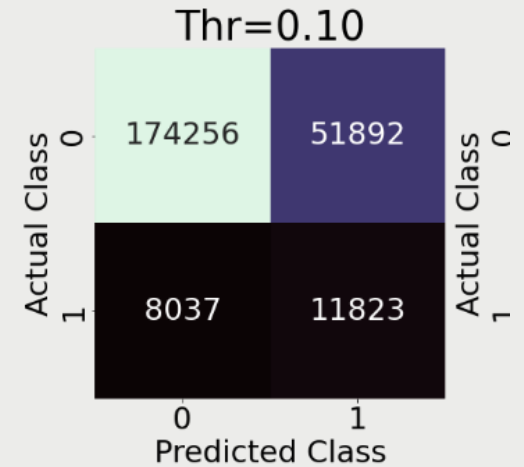
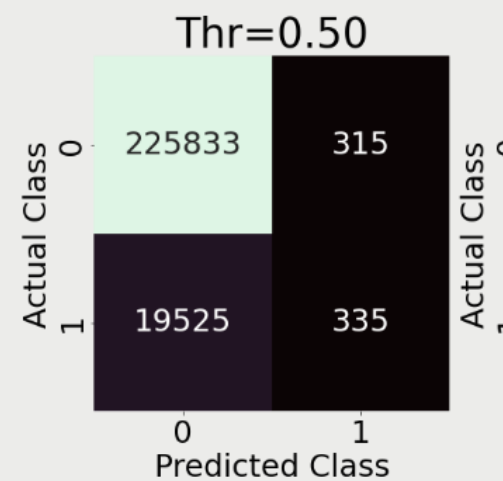
92 %





## Choix d'une métrique

Precision  
Recall  
F1 Score  
F bêta Score



### Dilemme

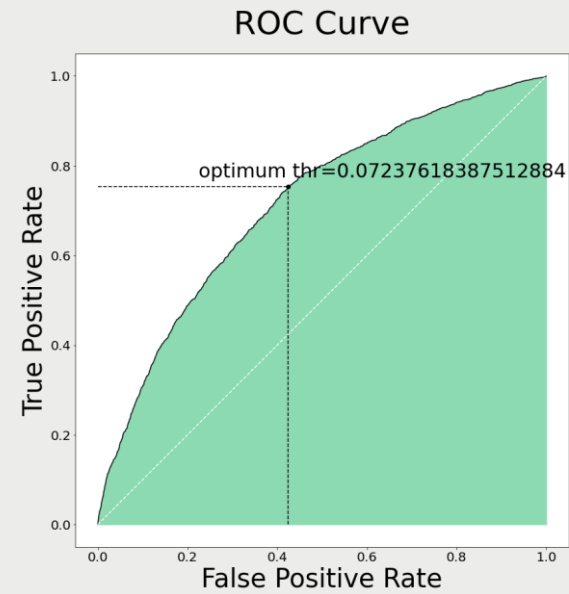
Accorder assez de crédits pour générer de l'argent

Avec assez de précaution pour ne trop en perdre



## Choix d'une métrique

ROC-AUC



Thr=0.10

Actual Class	0	1	Actual Class
	174256	51892	
1	8037	11823	1
Predicted Class			



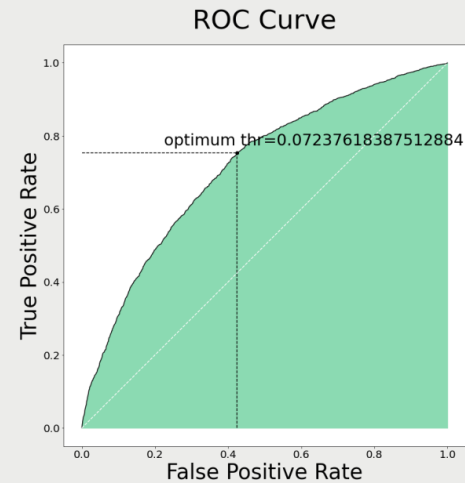


# Choix d'une métrique

## ROC-AUC

L'importance de la  
threshold

Predict\_proba()



		Thr=0.10		Thr=0.20	
Actual Class	0	174256	51892	212796	13352
	1	8037	11823	14437	5423
		Predicted Class		Predicted Class	

		Thr=0.30		Thr=0.40	
Actual Class	0	222136	4012	224915	1233
	1	17485	2375	18927	933
		Predicted Class		Predicted Class	



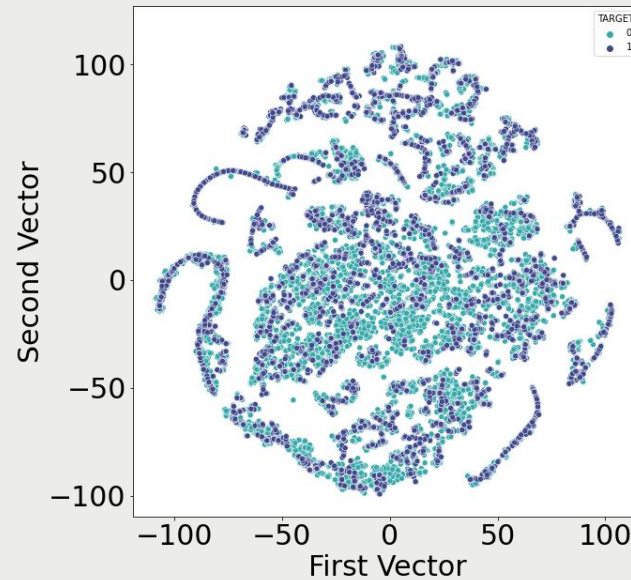
## Des cibles non balancées

Imblearn

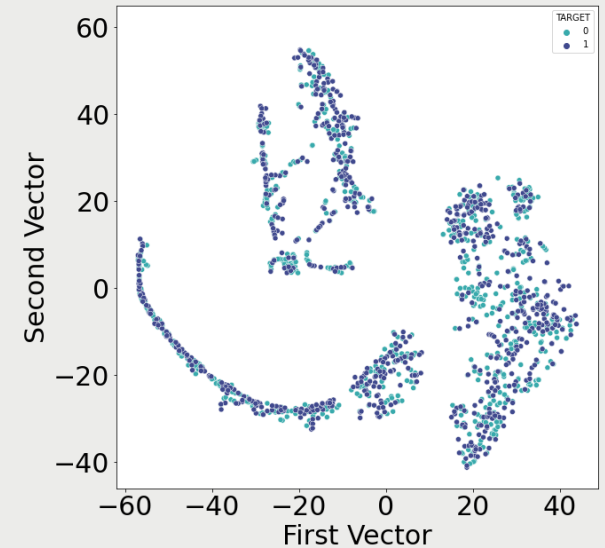
SMOTE,  
ADASYN,  
RandomUnderSampler,  
ClusterUnderSampler

Class weight

T-SNE SMOTE



T-SNE RandomUnderSampler



# Pipelines & Fine Tuning

Simple Imputer

One Hot Encoder

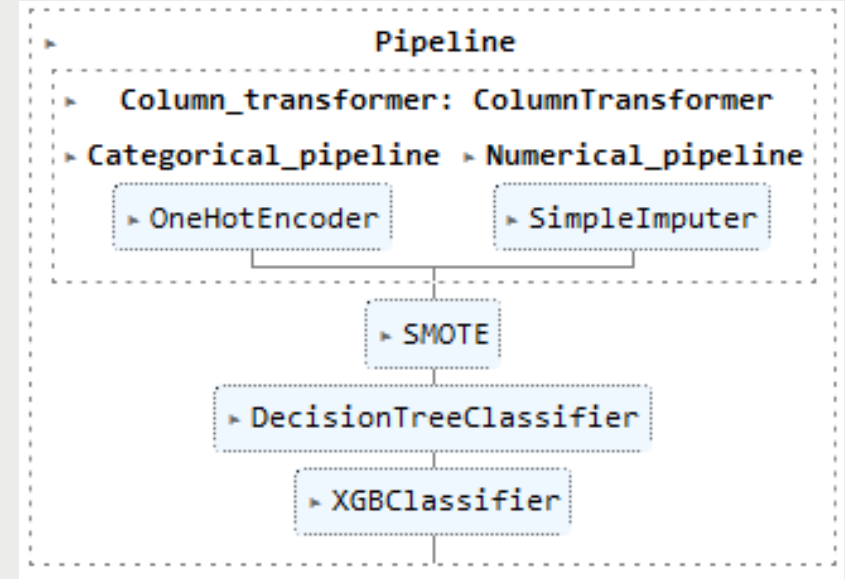
SelectFromModel

Standardisations

Sampling Method (Imblearn)

Réduction Dimensionnalité VS  
Interprétabilité

Pipeline Imblearn sklearn, Cross-validation (5folds), Stratification, GridSearch

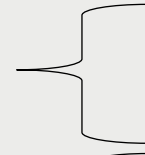






## Modèles Entraînés

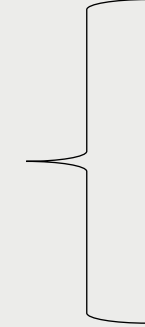
Linéaire



Régression Logistique

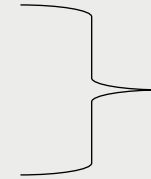
Random Forest

Arbres



XGBoost

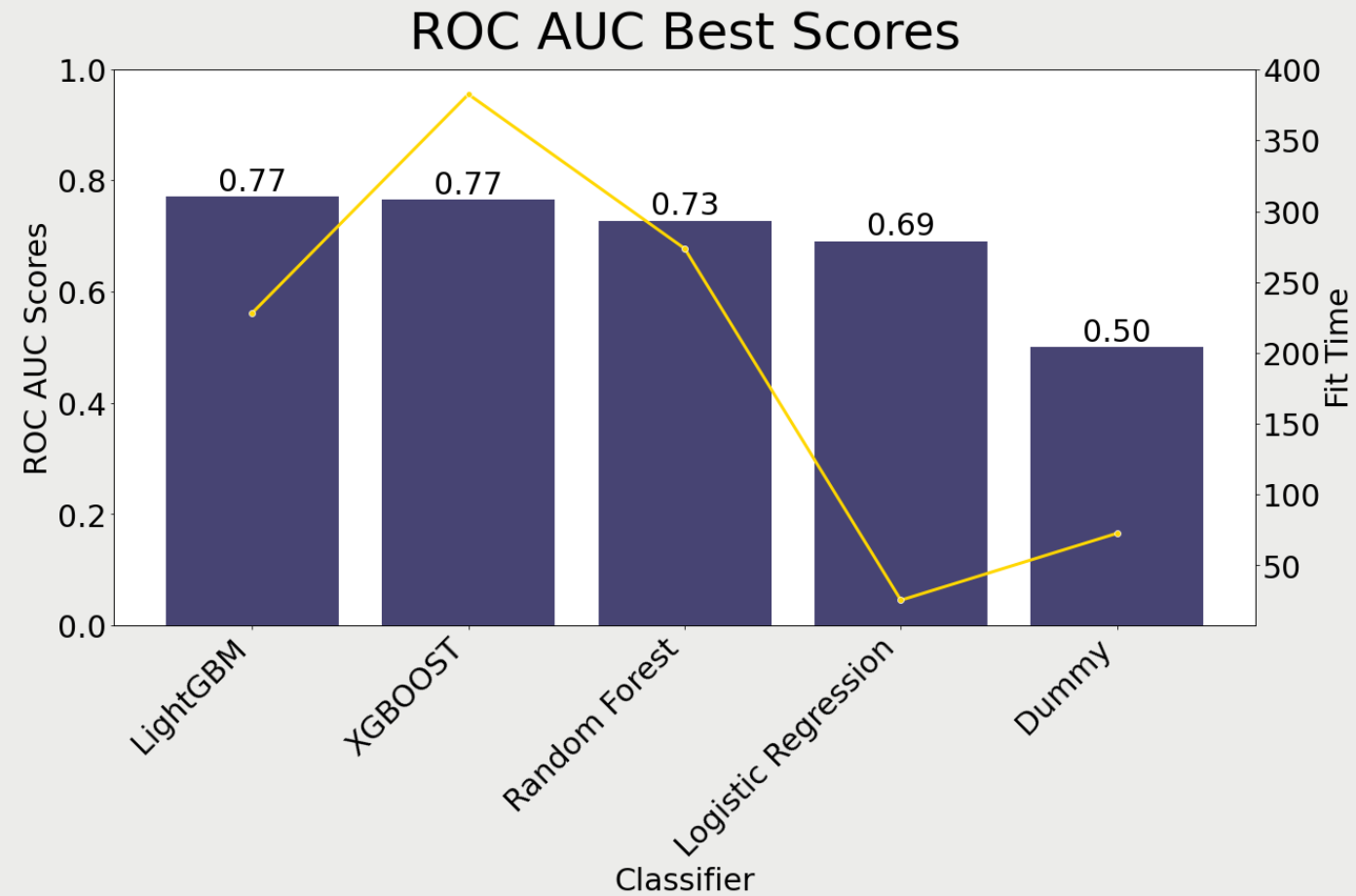
LightGBM



Gradient Boosting

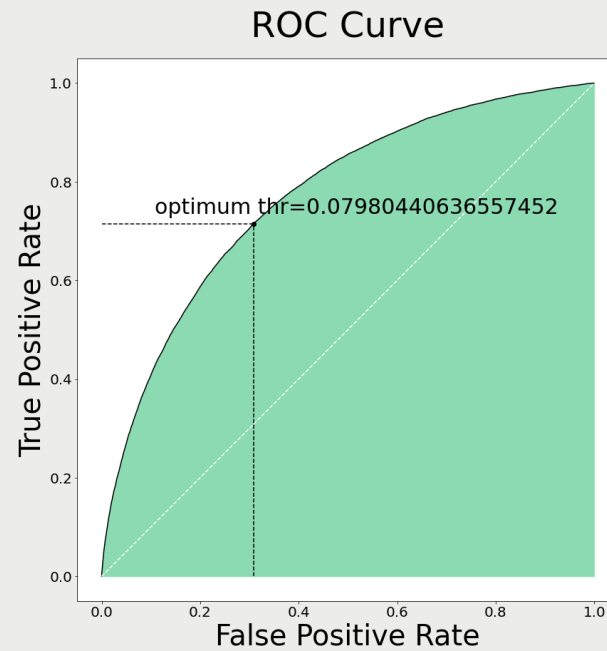


## Modèles Résultats





## Light GBM Global



Thr=0.08

Actual Class	Predicted Class	
	0	1
0	156490	69658
1	5663	14197

Threshold = 0,08





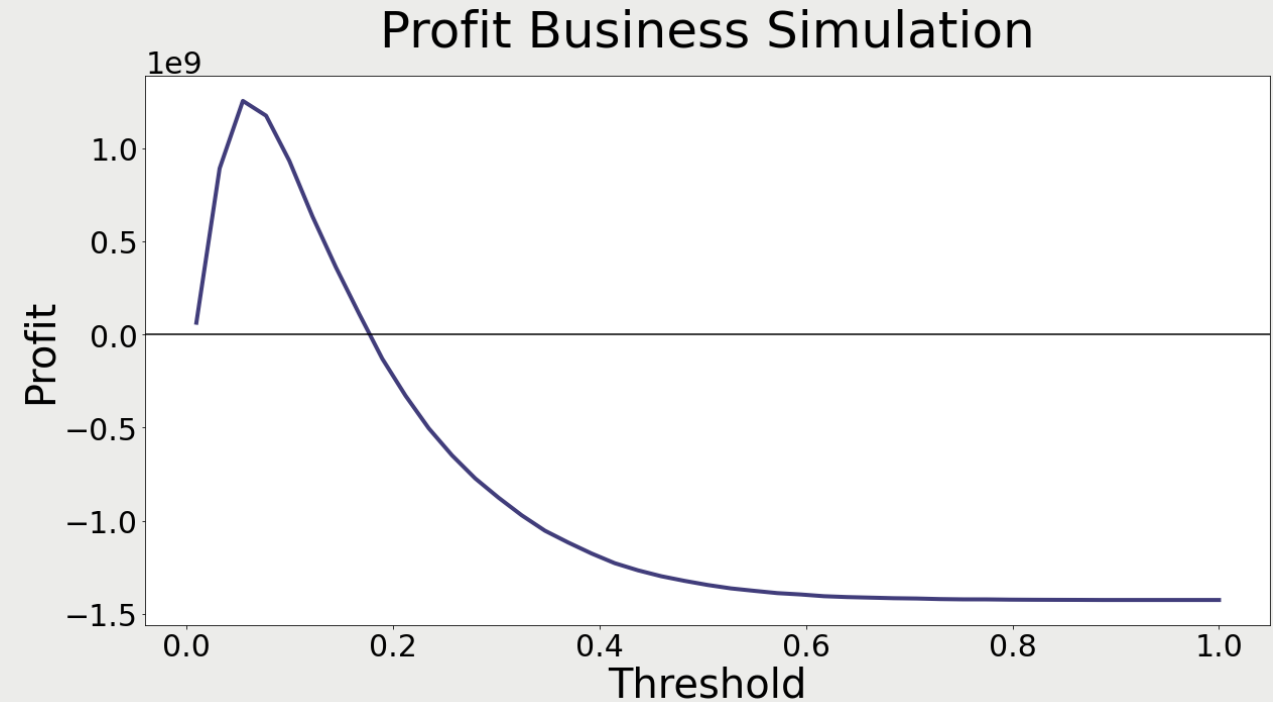
Hypothèses:

Gain : 1,03

Perte 0,5

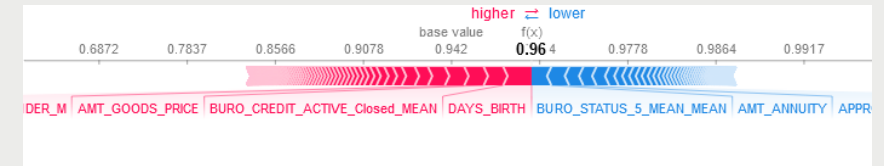
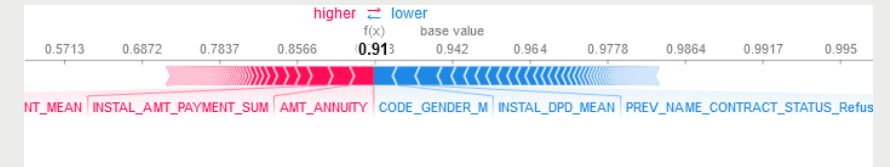
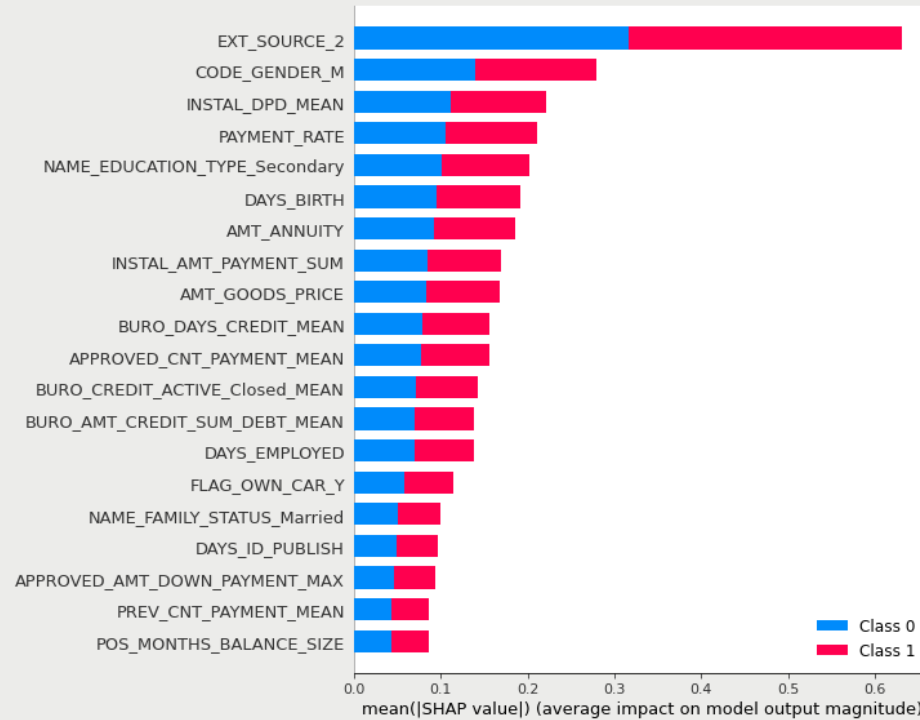
Threshold optimale  
correspond threshold  
optimale ROC-AUC

## Fonction Métier



# Light GBM Local

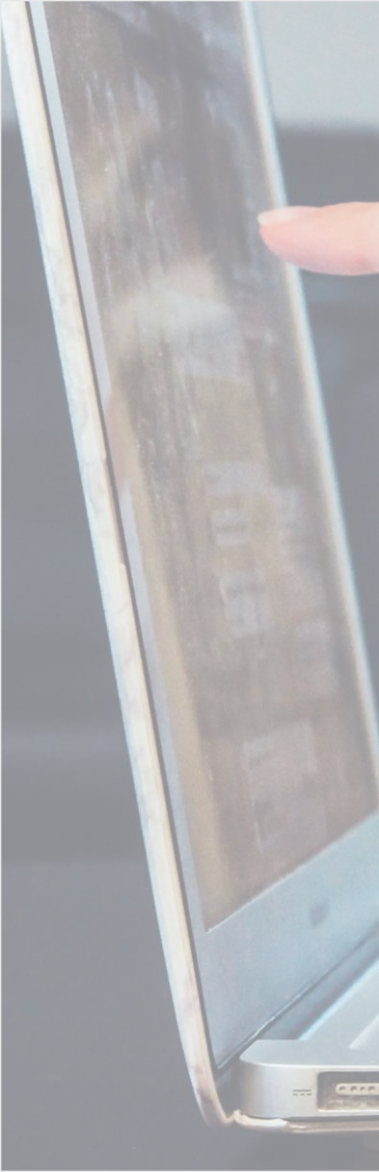
## Théorie des jeux coopératifs





### **III - Dashboard**





Flask

Clients.csv  
Modèle.pkl  
TreeExplainer

Routing  
Logiques métier

Protocole HTTP  
(GET, POST)

Json

## API

```
@app.route('/get_setup_infos')
def get_setup_infos():

    """Returns all features in the dataframe as well as all the categorical features only"""

    categories_list = ['EVERY_CLIENTS'] + dataframe.select_dtypes('object').columns.tolist()

    return {'all_features':dataframe.columns.tolist()[1:],
            'all_categories':categories_list}

@app.route('/get_prediction_proba/', methods=['POST'])
def get_prediction_proba():

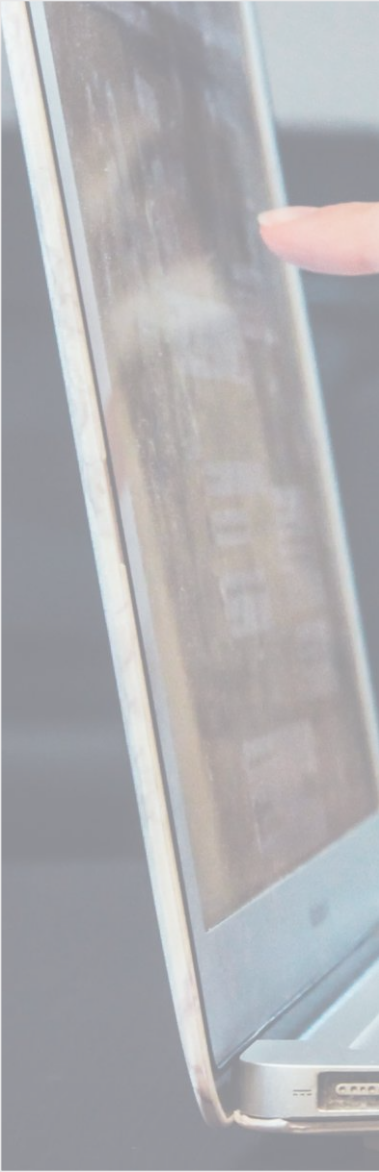
    """Returns the predicted proba of a given customer id"""

    request_data = request.get_json()

    # Retrieves the row of a user given an id
    user_data = dataframe[dataframe['SK_ID_CURR'] == int(request_data['customer_id'])].iloc[:,1:]

    if len(user_data) == 0:
        abort(404)
    else :
        # Uses the model to predict proba
        customer_prediction = lgbm_model.predict_proba(user_data).tolist()[0][0]

    return {'prediction': customer_prediction}
```



Streamlit

Composants faciles d'utilisation

Axé DataScience

Matplotlib compatible, Plotly

React

Limites : CSS, HTML compliqué  
(iframes), Markdown, attention  
injection HTML et JS

## Client

```
fig = px.box(  
    pd.DataFrame(  
        feature_group_value['values_list'],  
        columns=['SK_ID_CURR',selected_feature]),  
    y=selected_feature,  
    hover_data=['SK_ID_CURR',selected_feature],  
    points="all",  
    labels={  
        "variable": re.sub('_', ' ', selected_feature.lower())  
    },  
    title='{} representation inside {} category = {}'.format(re.sub('_', ' ', selected_feature.lower()),  
                                                             re.sub('_', ' ', selected_category.lower()),  
                                                             category_customer_value['feature_customer_value'])  
  
fig.update_xaxes(tickvals=[""])  
fig.add_hline(y=feature_customer_value['feature_customer_value'],  
              annotation_text="You are here")  
st.plotly_chart(fig)
```

# Déploiement dans le cloud

En local Environnement Virtuels  
(Venv)

Pip freeze > requirements.txt

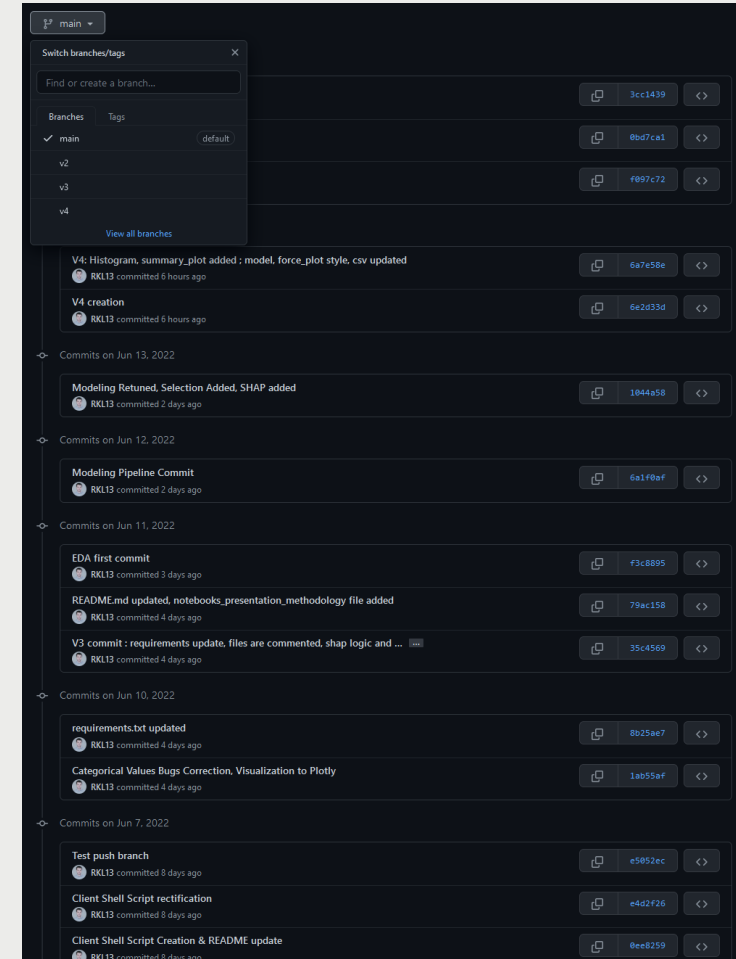
Versioning (4 Versions, 4 Branches)  
Git, GitHub

2 Serveurs AWS

Linux distribution Ubuntu


Tmux

Script Shell pour le déploiement





# UX



## Dashboard

Enter a customer ID

You can try 100002 or 100003

Choose a feature to explore:

NAME\_CONTRACT\_TYPE ▾

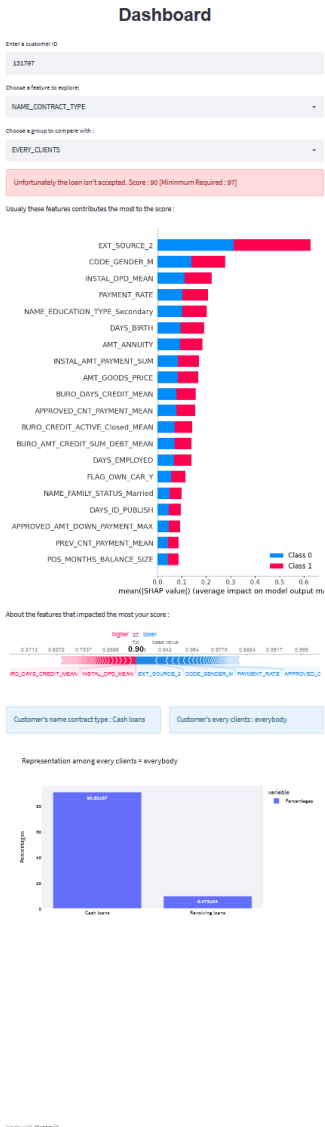
Choose a group to compare with :

EVERY\_CLIENTS ▾

The selected ID isn't correct.

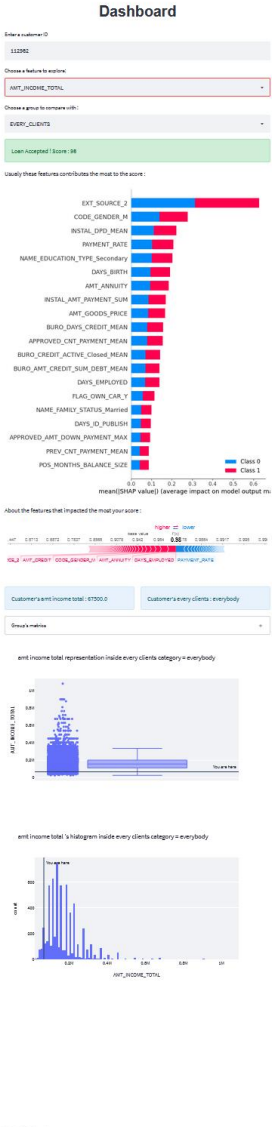
Made with Streamlit

<http://18.233.144.105:8501/>



UX

<http://18.233.144.105:8501/>



## IV - Limites et Améliorations



# Limites et améliorations

Plus de samples de clients non solvables

Voting Classifier

Base de données, CRUD, Simulations

Docker

Score, Vulgarisation plus poussée, customer centric, lean management, feedbacks

A pink piggy bank is the central focus, standing on a surface covered with numerous gold coins. The piggy bank has a coin slot on its back and two small circular holes on its snout. The background is a soft, out-of-focus yellow and white. The text "Merci pour votre attention !" is superimposed over the piggy bank's body.

**Merci pour votre attention !**