

The background is a blurred image of a document with a pen. A large white circle is centered on the page. To the left of the circle, there are several teal-colored dashed lines. The text 'Prêt à dépenser' is written in a large, bold, black font inside the circle. Below it, the text « Modèle de scoring » is written in a smaller, black font. In the bottom right corner of the circle, there is a small circular logo with the text 'Prêt à dépenser' inside.

Prêt à dépenser

« Modèle de scoring »

Prêt à dépenser

Sommaire



Contexte



Présentation du jeu de données



Traitement du jeu de données



Modélisation



Explication du modèle



Conclusion



Contexte



Contexte



CRÉDIT À LA
CONSOMMATION



FACILEMENT
INTERPRÉTABLE



IMPORTANCE DES
VARIABLES

Présentation du jeu de données

```
error  
ject  
_x":  
rue  
alse  
alse  
OR_y":  
alse  
rue  
alse  
OR_z":  
alse  
alse  
rue
```

```
d -add
```

```
s.active  
odifier  
0  
ted_object  
me].select
```

```
exactly
```

```
S
```

```
selected  
r_x"
```

```
is not
```

Analyse de la forme

307.511
Lignes et 122
colonnes

Répartition
« Target »
déséquilibrée

	Nombre de valeur	Pourcentage
0	282686.0	91.93
1	24825.0	8.07

Analyse de la forme

307.511
Lignes et 122
colonnes

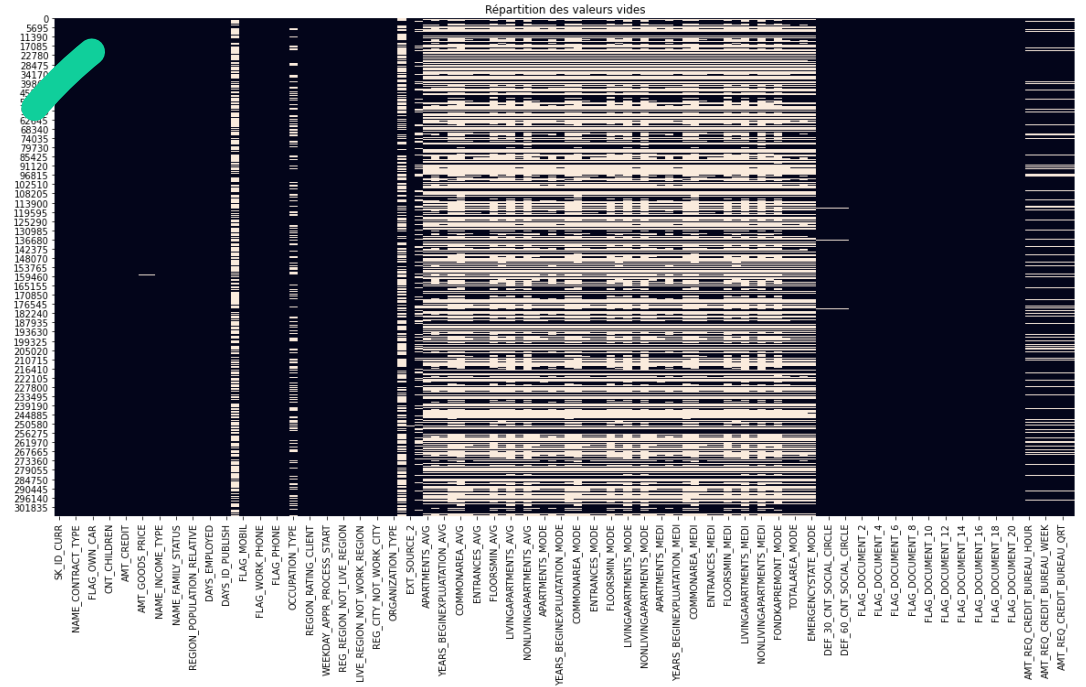
Répartition
« Target »
déséquilibrée

	Nombre de valeur	Pourcentage
remboursé	282686.0	91.93
non-remboursé	24825.0	8.07

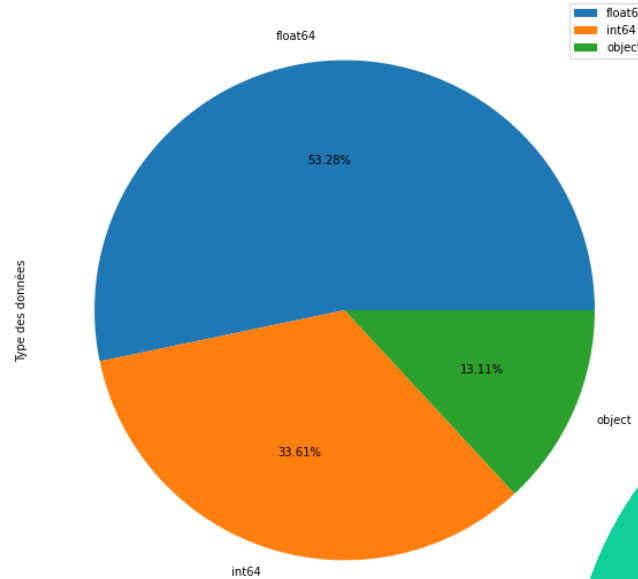
1

Répartitions des valeurs vides

Répartition des types de données



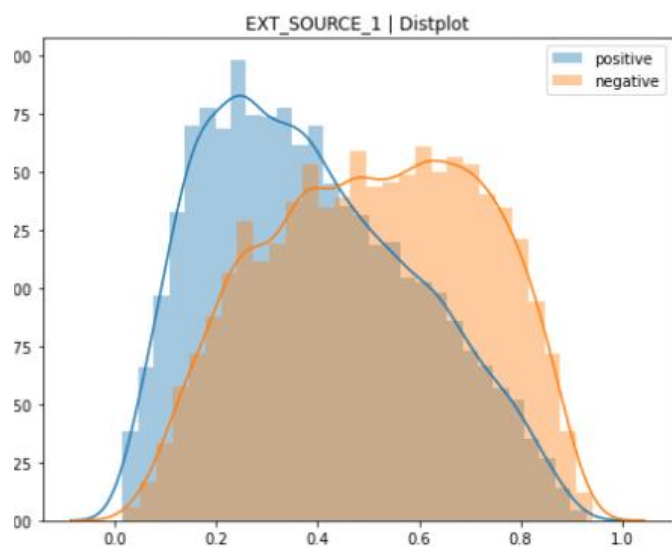
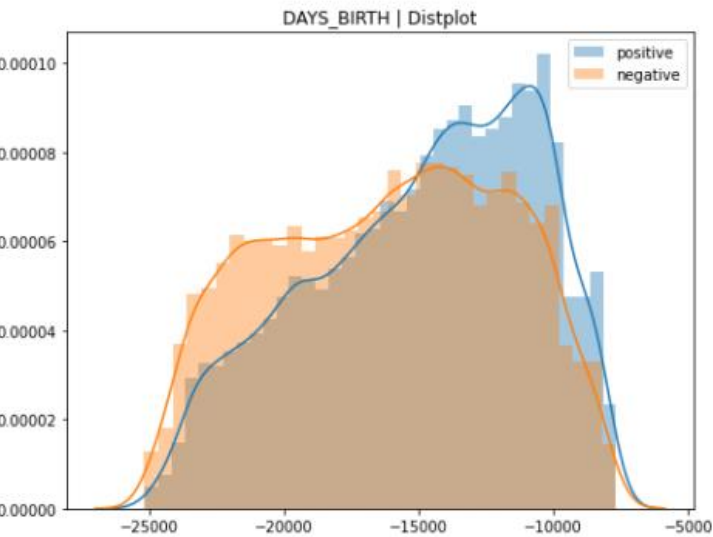
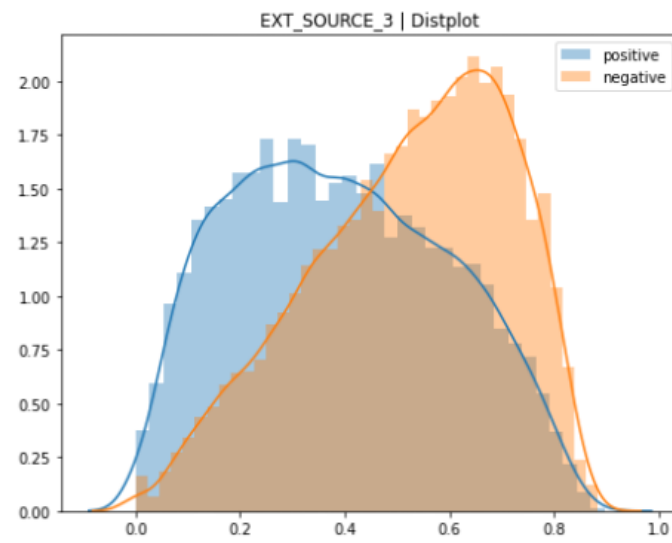
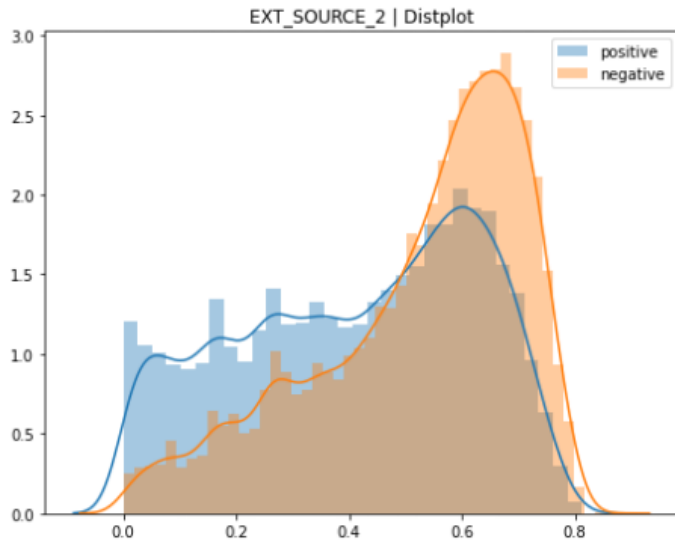
Répartition des types dans le jeu de données.



Analyse de fond



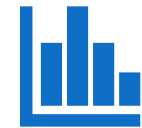
Valeurs numériques



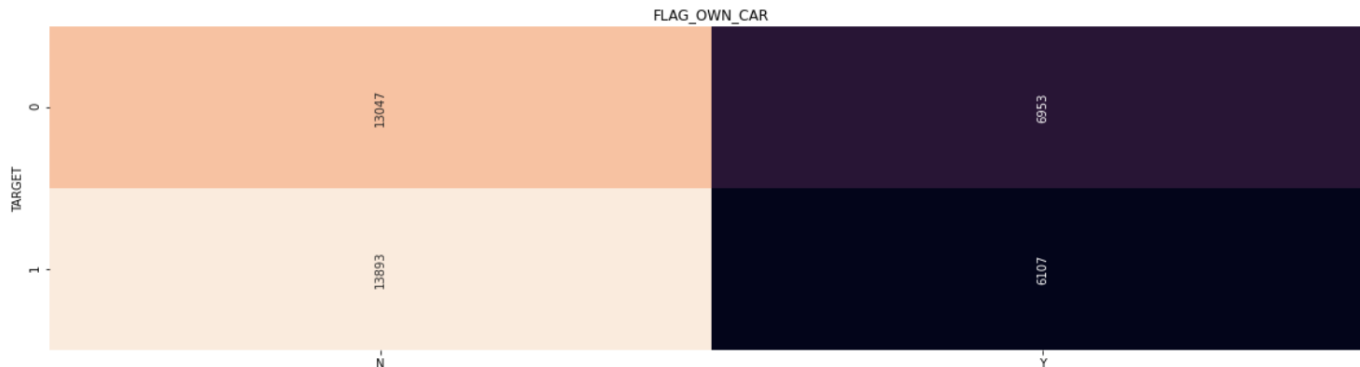
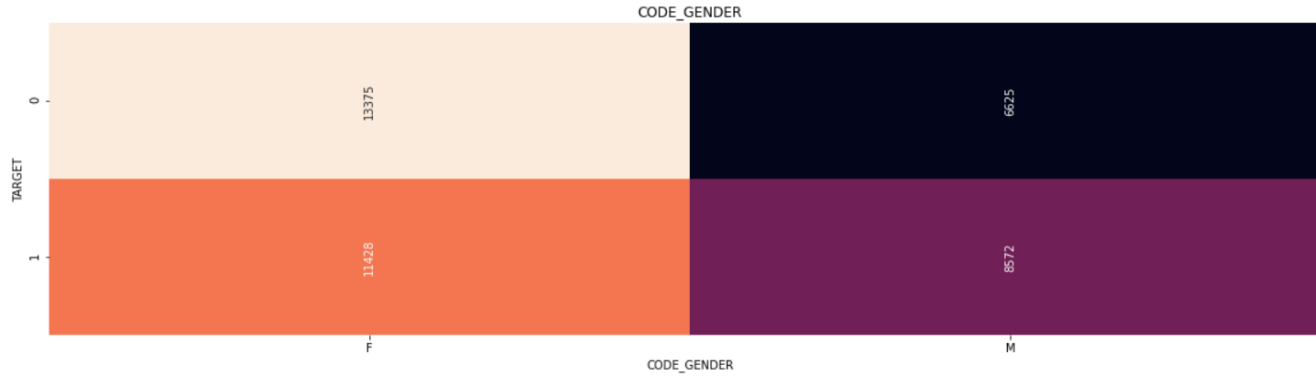
Analyse de fond



Valeurs
numériques



Valeurs
catégorielles

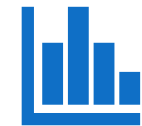




Analyse de fond



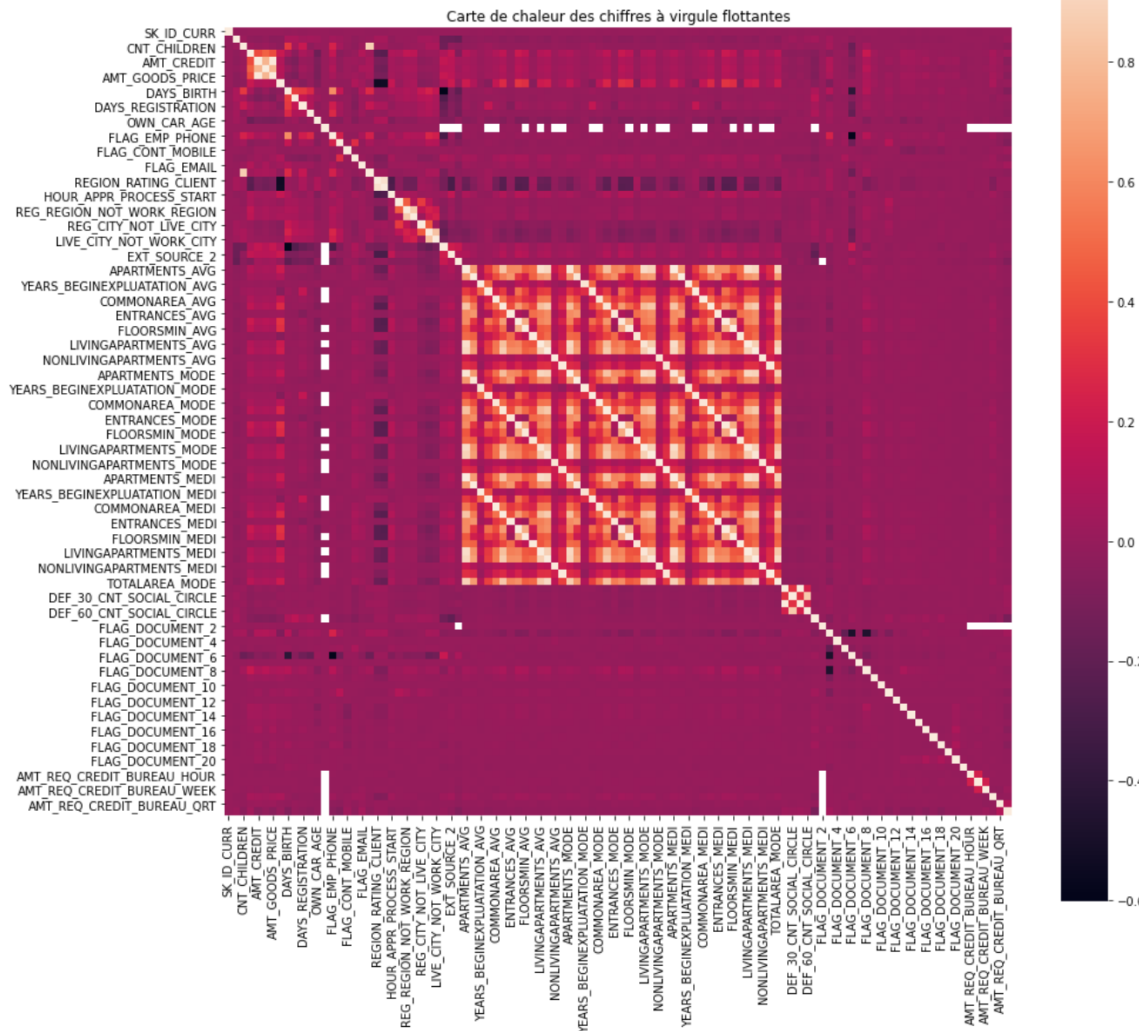
Valeurs
numériques



Valeurs
catégorielles



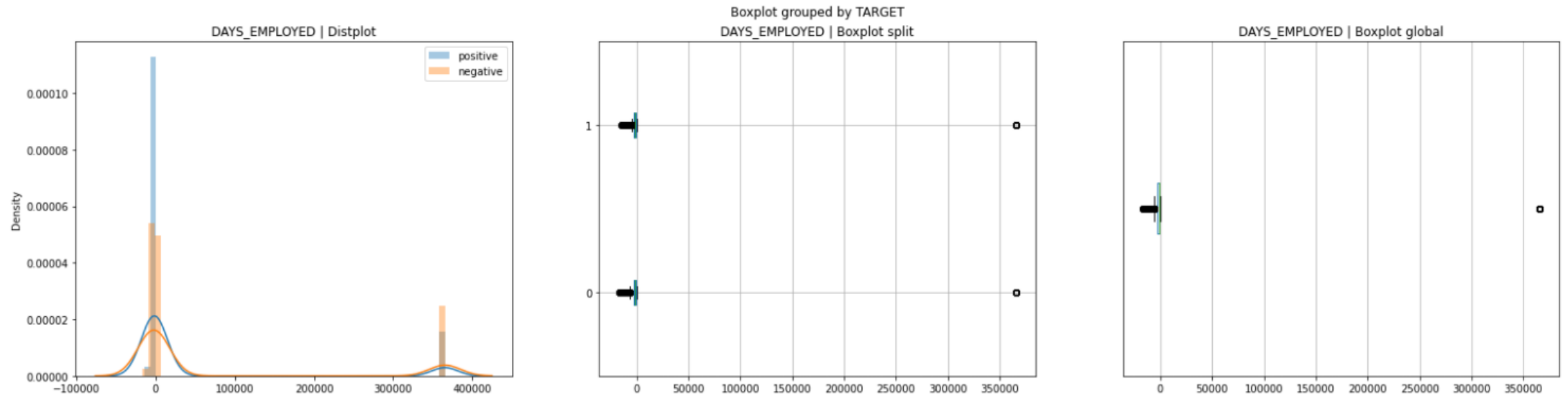
Corrélation valeurs
numériques



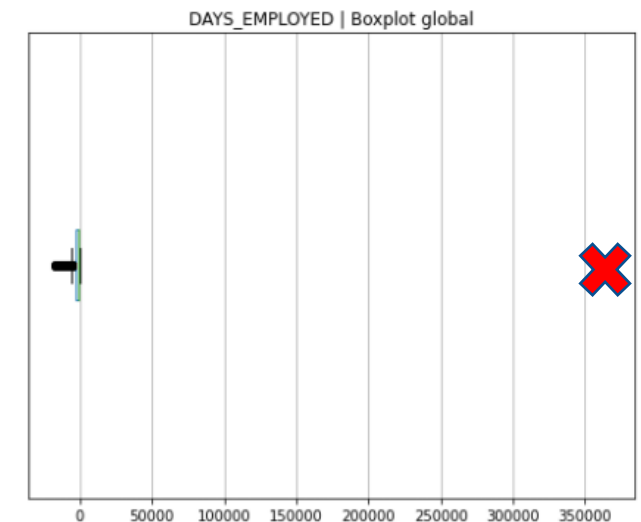
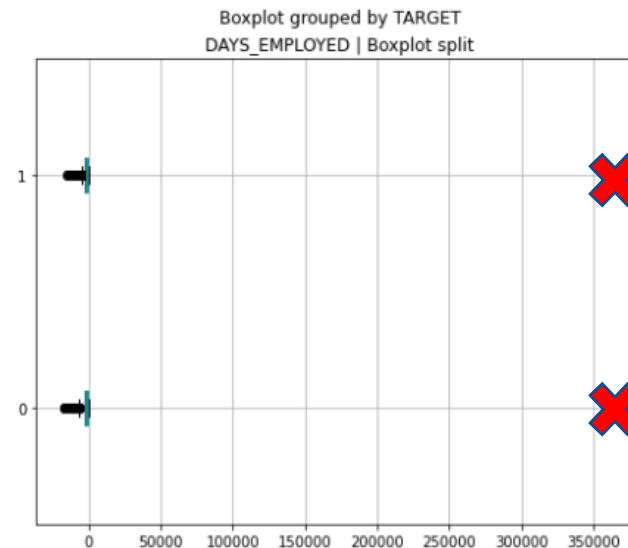
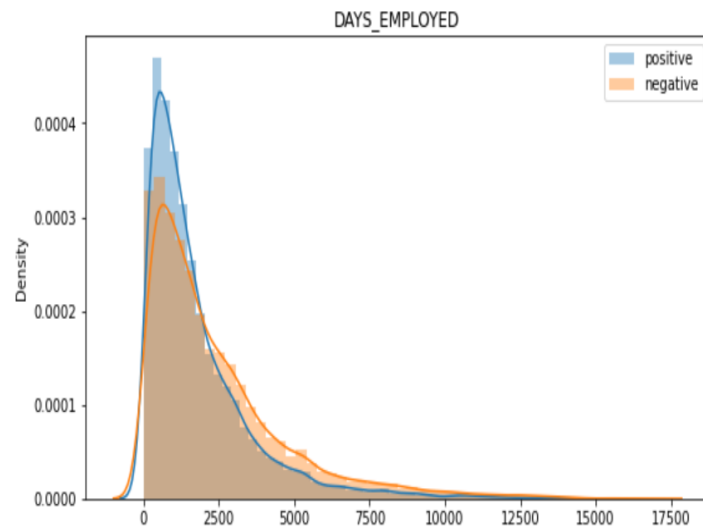
Traitement du jeu de données



Valeurs aberrantes



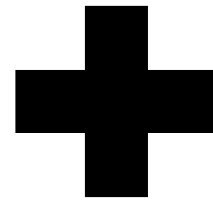
Valeurs aberrantes



Imputation

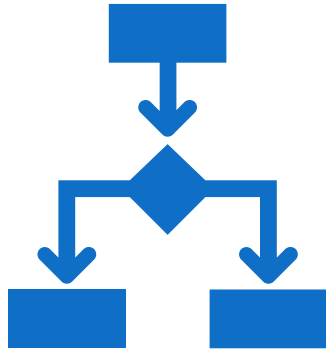


Imputation
Médiane



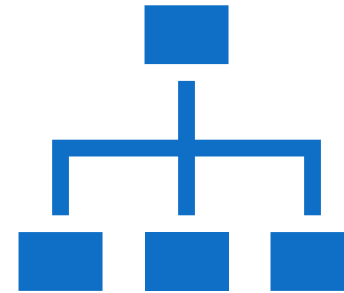
Reste à 0

Données catégorielles



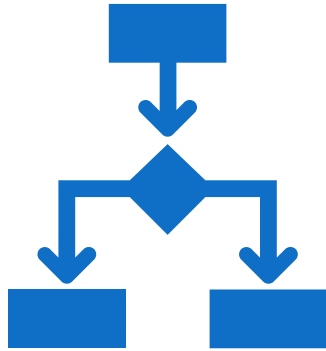
LabelEncoder (unique
valeurs ≤ 2)

/	ANIMAL
1	CHIEN
2	CHAT



OneHotEncoder

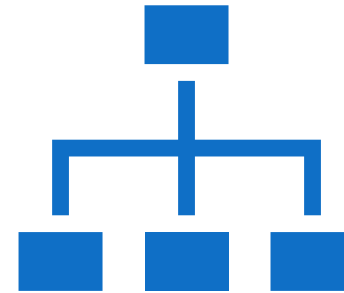
Données catégorielles



LabelEncoder (unique
valeurs ≤ 2)

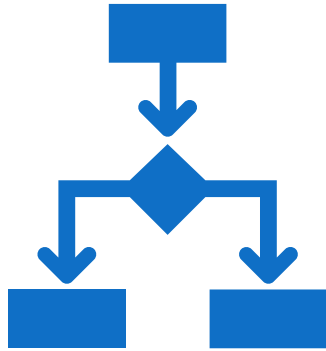
/	ANIMAL
1	0
2	1

- CHIEN = 0
- CHAT = 1



OneHotEncoder

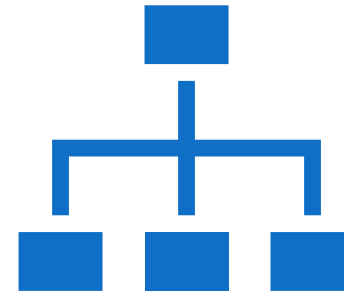
Données catégorielles



LabelEncoder (unique
valeurs ≤ 2)

/	ANIMAL
1	0
2	1

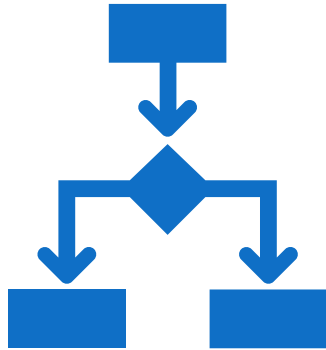
- CHIEN = 0
- CHAT = 1



OneHotEncoder

	ANIMAL_
1	CHIEN
2	CHAT

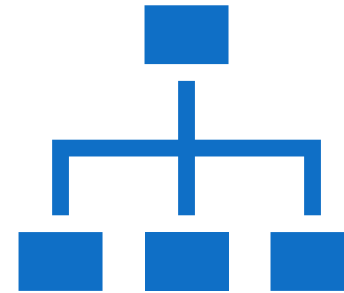
Données catégorielles



LabelEncoder (unique valeurs ≤ 2)

/	ANIMAL
1	0
2	1

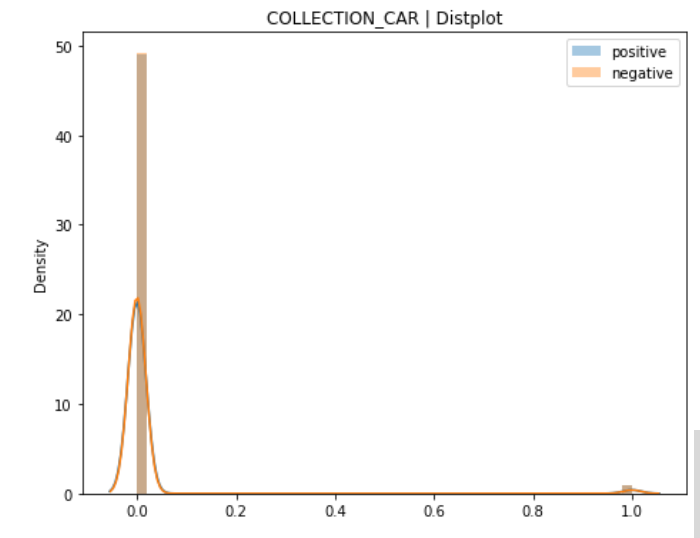
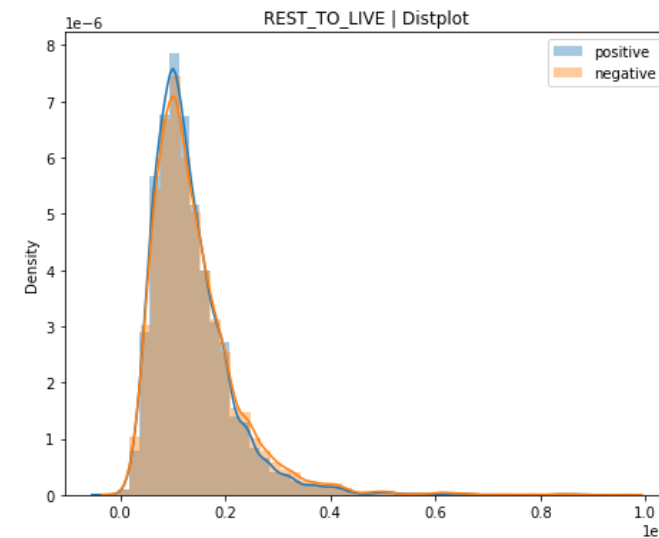
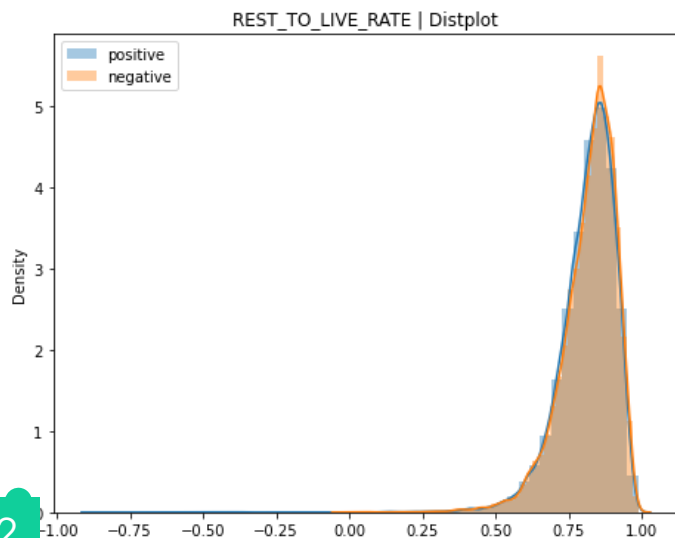
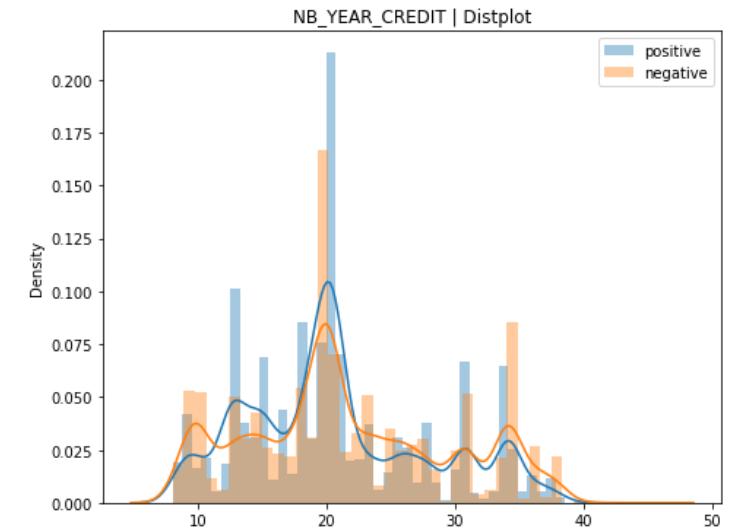
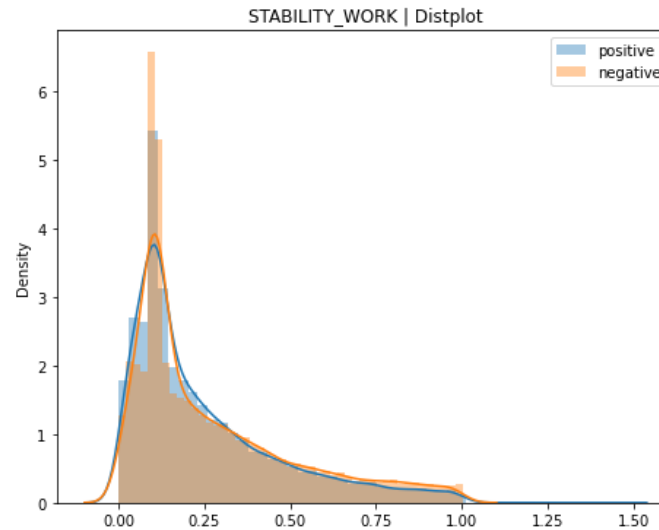
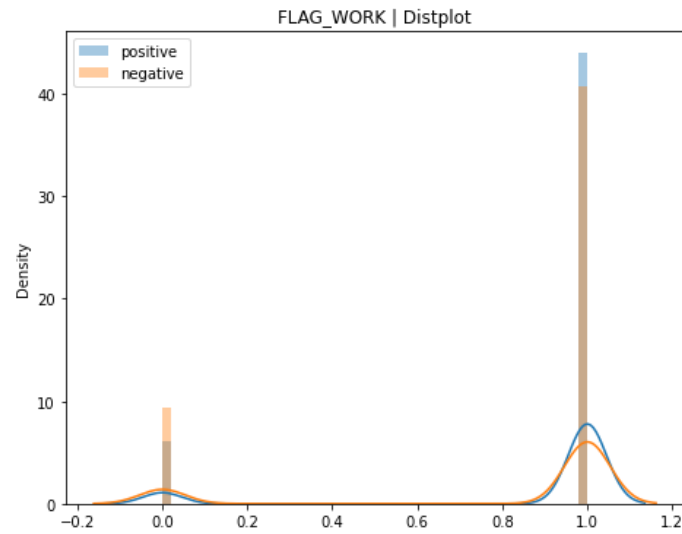
- CHIEN = 0
- CHAT = 1



OneHotEncoder

	ANIMAL_CHIEN	ANIMAL_CHAT
1	1	0
2	0	1

Création de nouvelles colonnes



Sélections des données



MÉTIER



BORUTA



STATISTIQUE



Modélisation



Métrique

- F_β -score (β à 2)
- Recall
- Precision



Choix des modèles



SupportVectorClassifier

LogisticRegression

KNeighborsClassifier

RandomForest

AdaBoost

XGBoost

Tests itératifs

GridSearchCV

~~SupportVectorClassifier~~

LogisticRegression

~~KNeighborsClassifier~~

RandomForest

~~AdaBoost~~

XGBoost

Dataset 350.000 lignes et 67 colonnes...

	RandomForest_no_prep	AdaBoost_no_prep	LogisticRegression_no_prep	KNN_no_prep	Xgboost_without_prep	Naive
recall	0.627	0.0	0.666	0.049	0.636	0.498
precision (label 1)	0.171	0.0	0.157	0.169	0.181	0.081
accuracy	0.724	0.919	0.685	0.904	0.738	0.5
fbeta_2-score	0.409	0.0	0.404	0.057	0.423	0.245
Temps_d'entrainement	6.607m	5.827m	0.403m	15.561m	11.998m	None


Optimisation du modèle

RandomForest

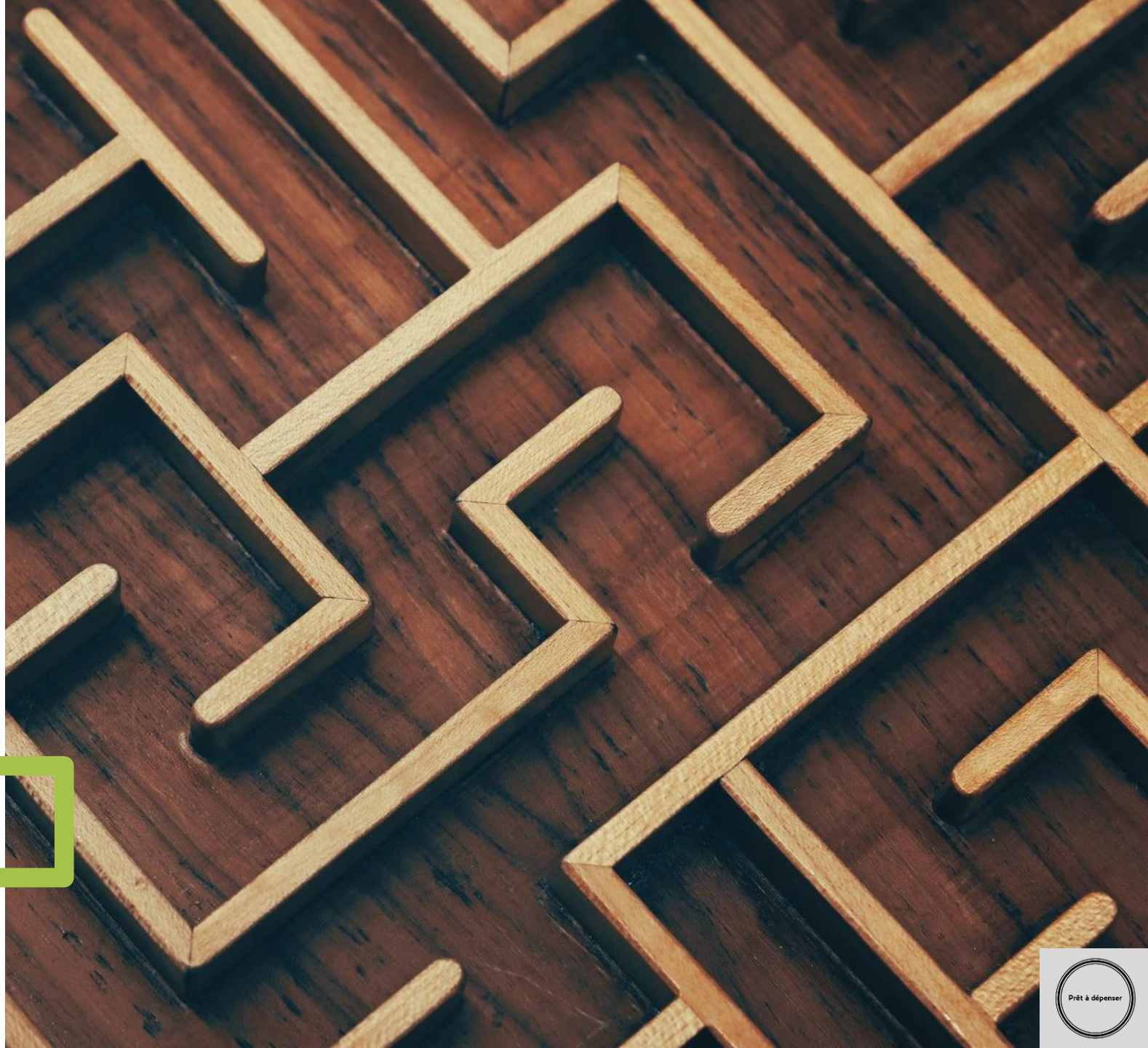


- Hyperparamètres finaux :
 - N_estimators : 250
 - Criterion : « entropy »
 - Max_depth : 10
 - Class_weight : balanced

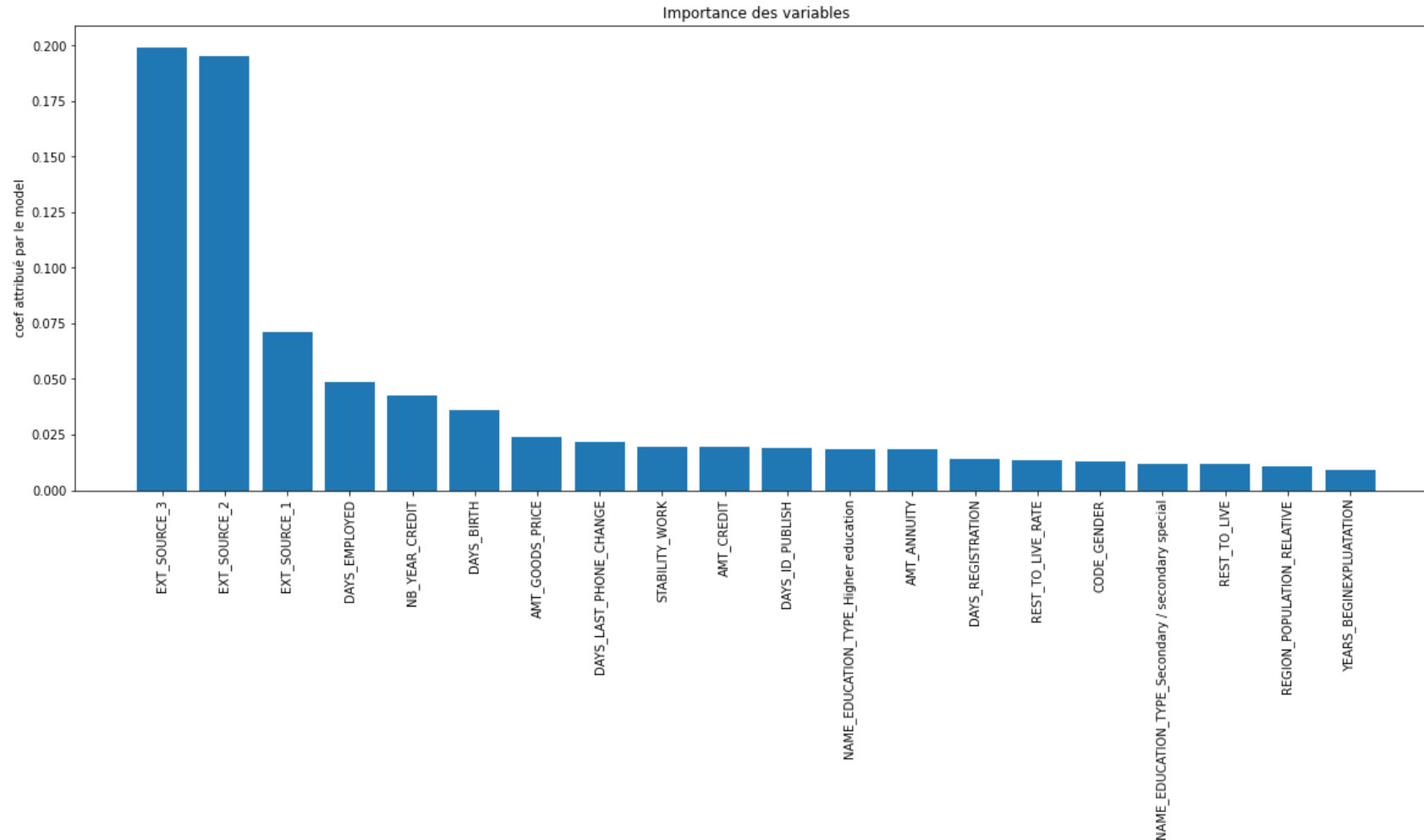




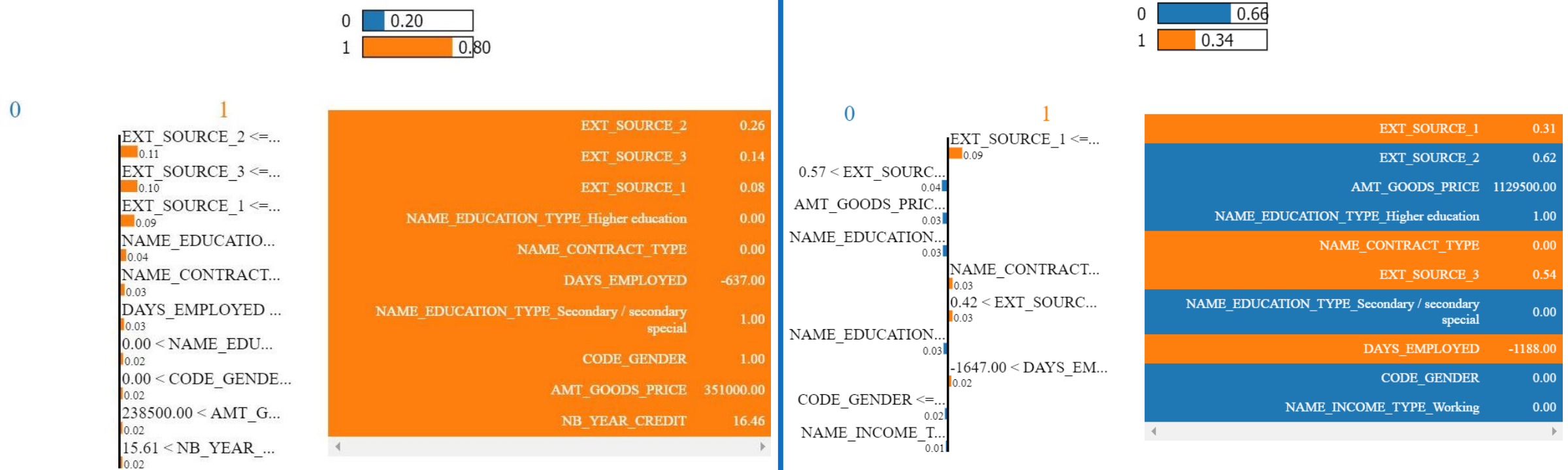
Explication du modèle



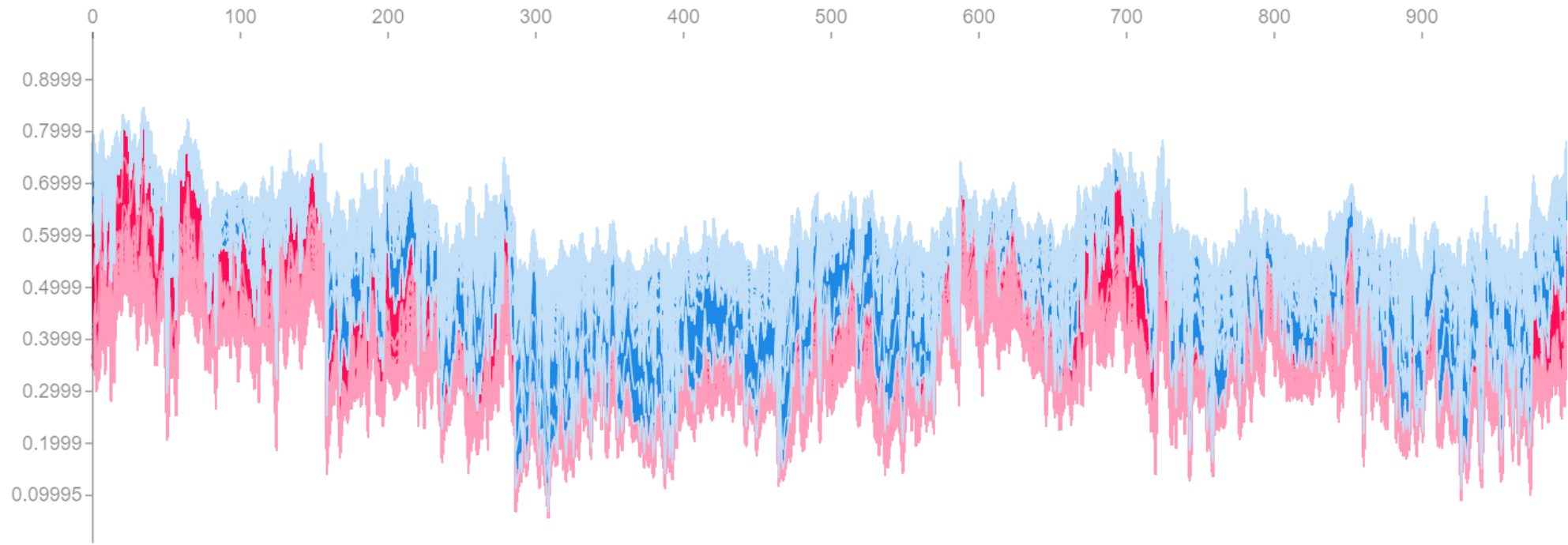
Explication simple



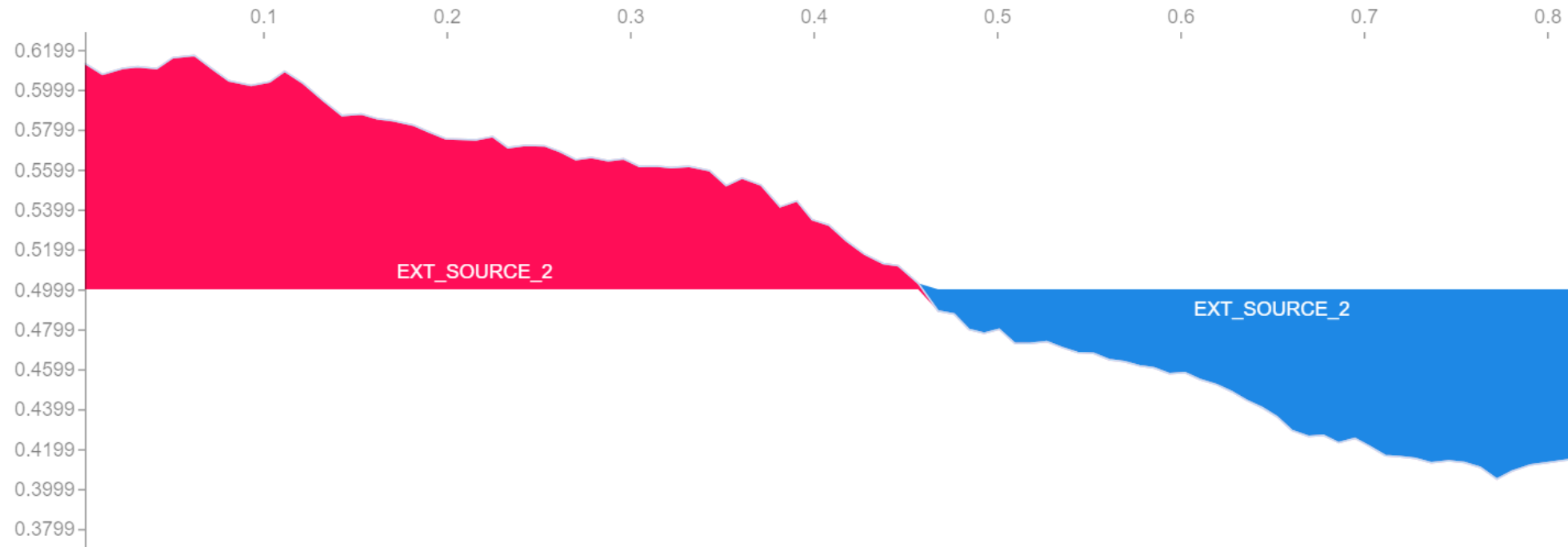
Explication par LIME



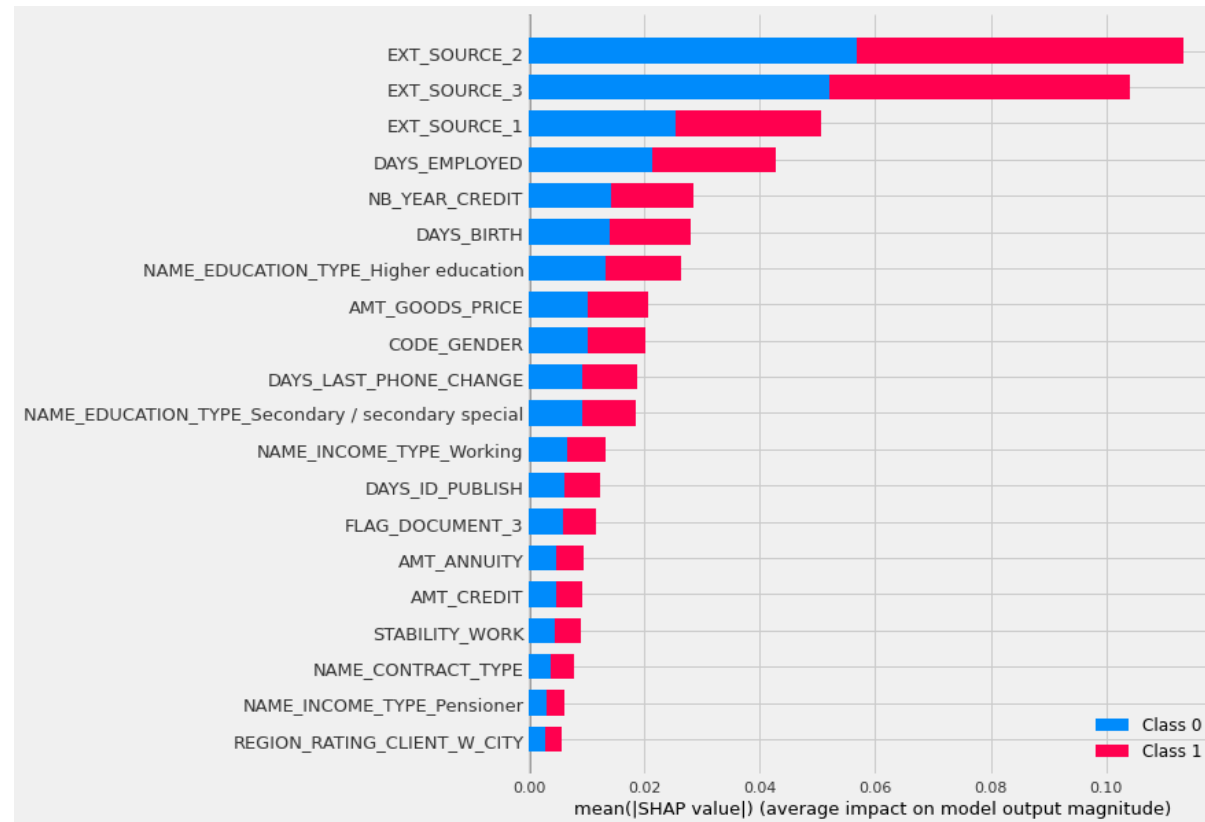
Explication par SHAP



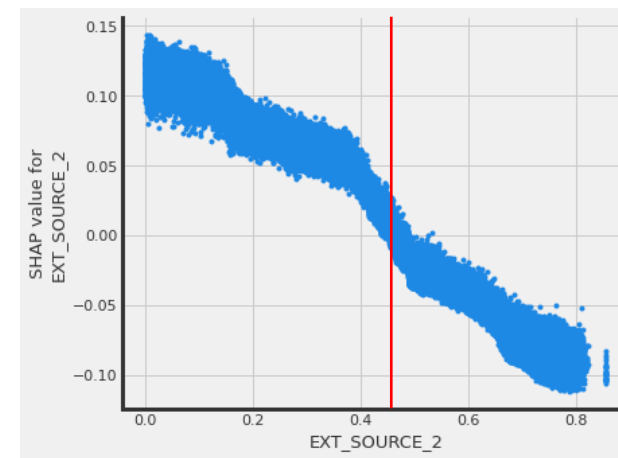
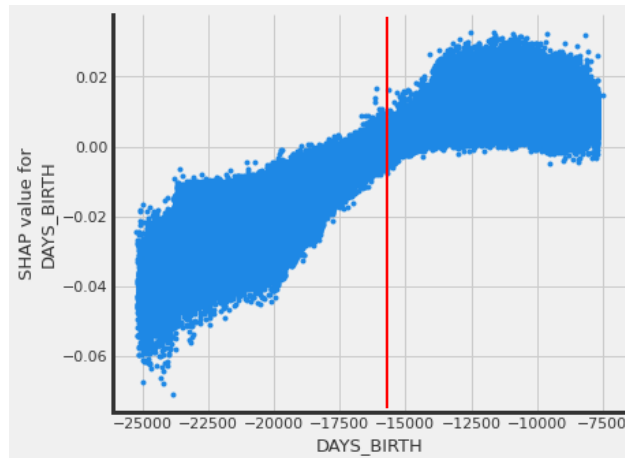
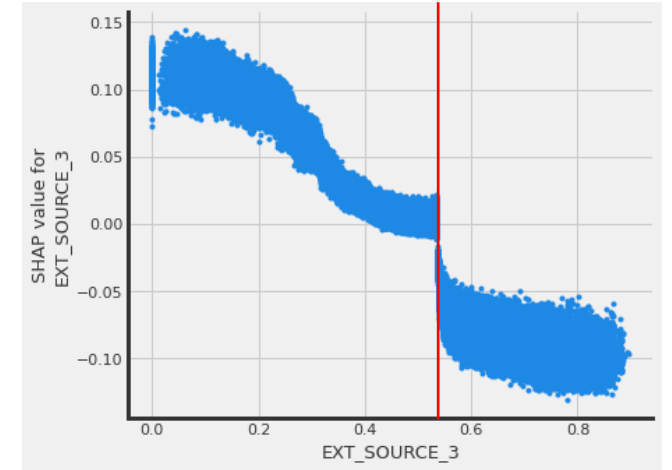
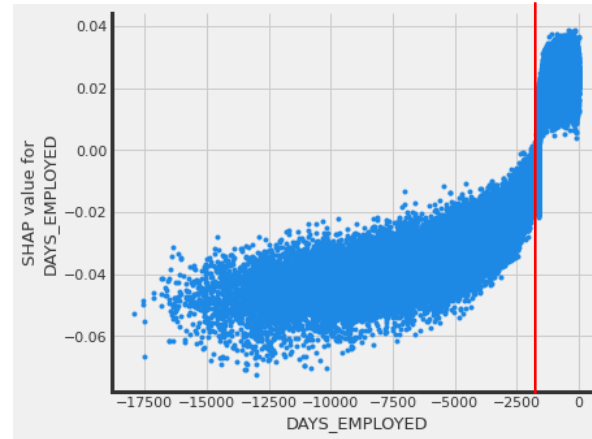
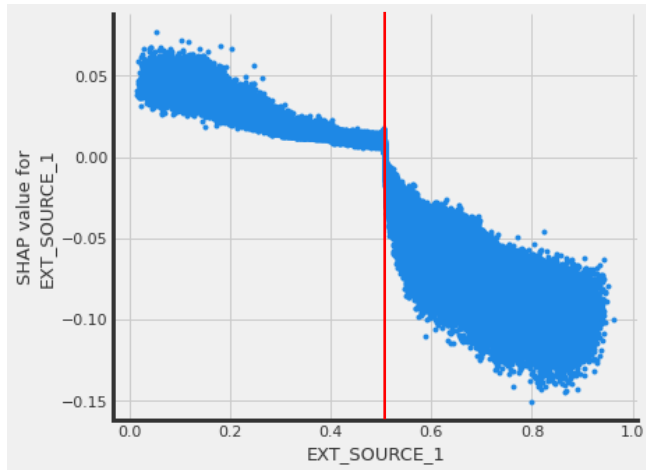
Explication par SHAP



Explication par SHAP



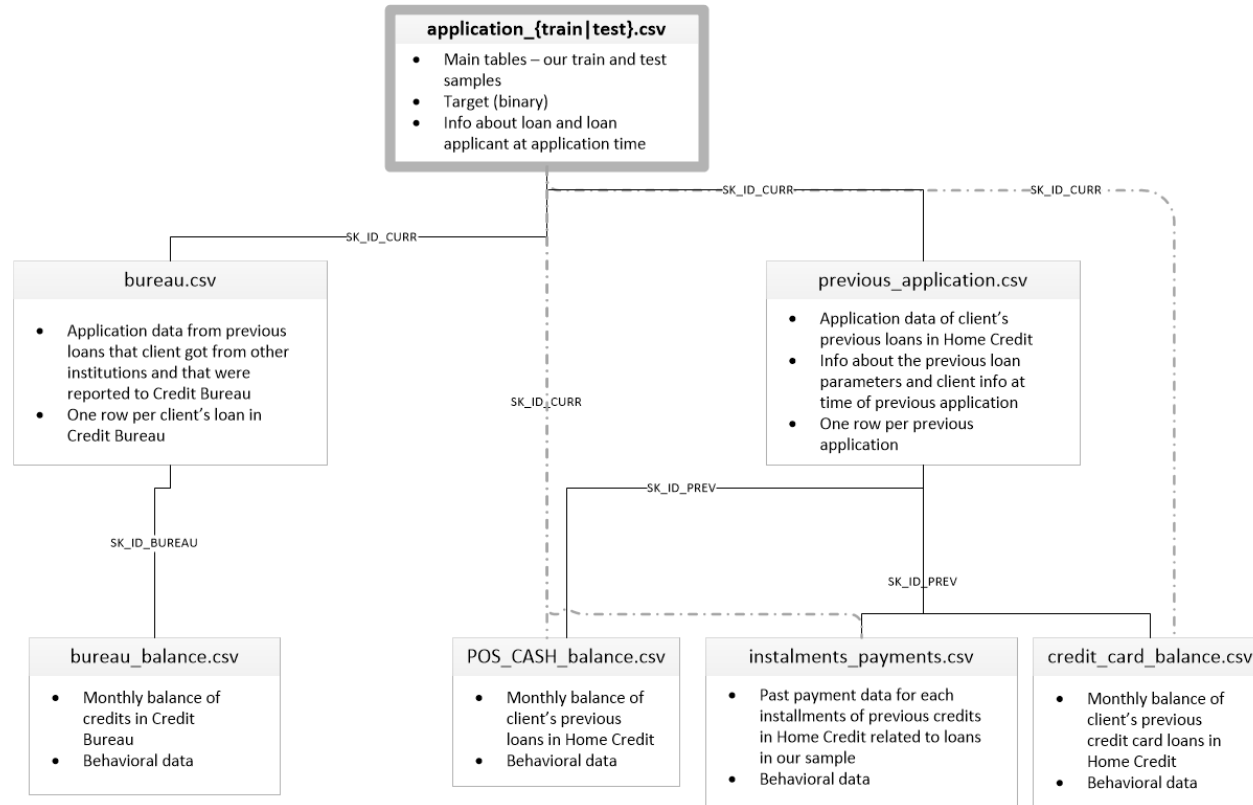
Explication par SHAP



Conclusion



Conclusion



Score 0,408



Prendre tous les fichiers



Premier projet ML



**Merci de
votre
attention,
avez-vous des
questions ?**

