



Prêt à dépenser

«Modèle de scoring»



By Tyson JOHN

Sommaire



Contexte



Présentation du jeu de données



Traitement du jeu de données



Modélisation

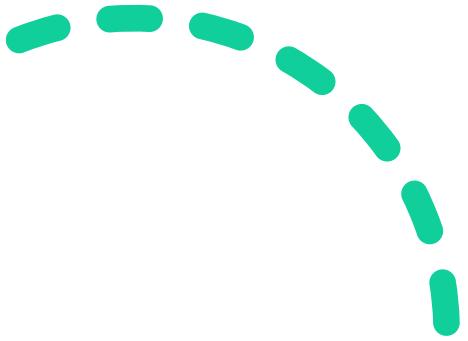


Explication du modèle

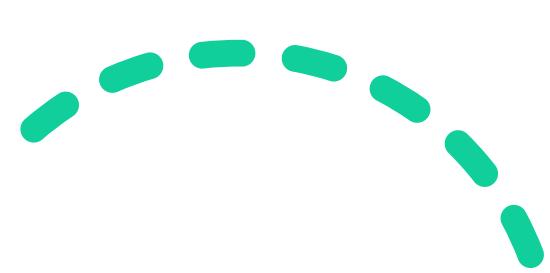


Conclusion

Contexte



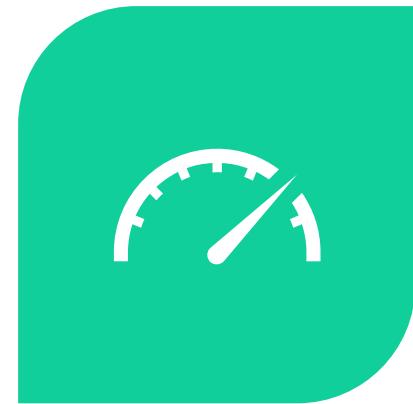
Contexte



CRÉDIT À LA
CONSOMMATION

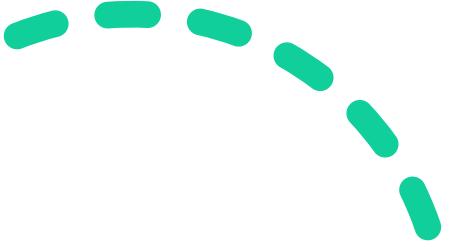


FACILEMENT
INTERPRÉTABLE

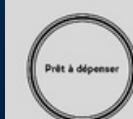


IMPORTANCE DES
VARIABLES

Présentation du jeu de données



```
rror  
ject  
  
_X":  
true  
false  
false  
OR_Y":  
false  
true  
false  
OR_Z":  
false  
false  
true  
  
d -add  
  
ts.active  
modifier  
0  
ited_obj  
ame].sel  
: exact  
S -  
er  
or_X"  
  
is not
```

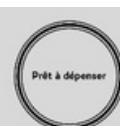


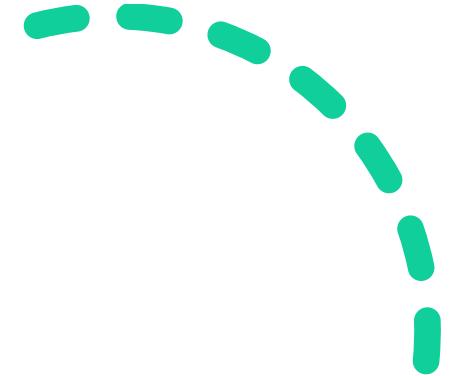
Analyse de la forme

307.511
Lignes et 122
colonnes

Répartition
«Target»
déséquilibrée

	Nombre de valeur	Pourcentage
0	282686.0	91.93
1	24825.0	8.07

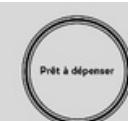




307.511
Lignes et 122
colonnes

Répartition
«Target»
déséquilibrée

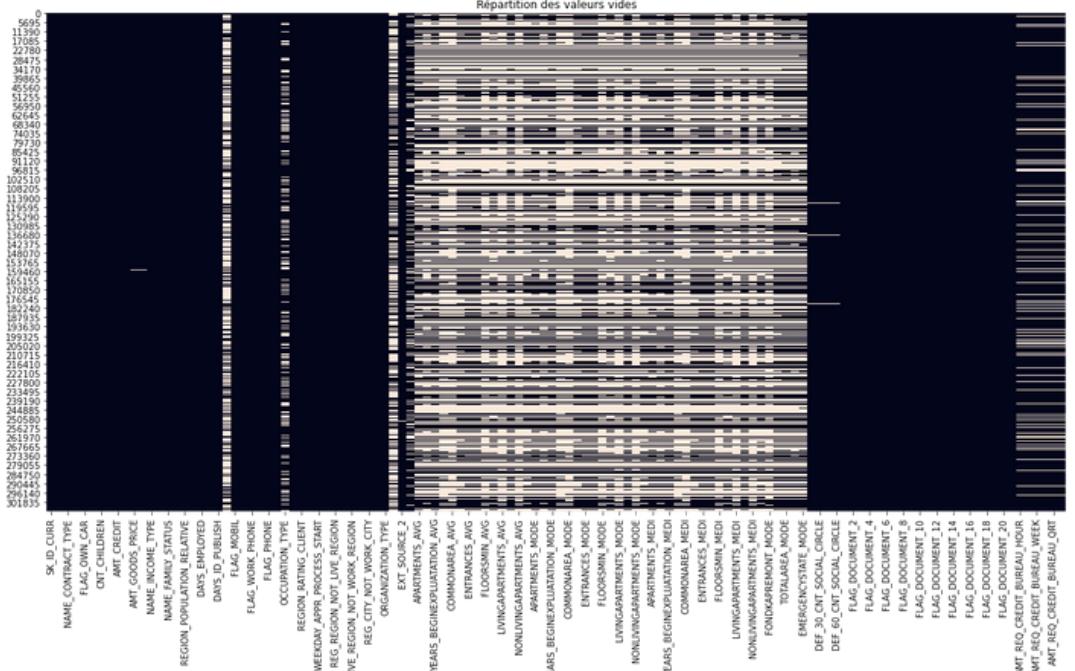
	Nombre de valeur	Pourcentage
remboursé	282686.0	91.93
non-remboursé	24825.0	8.07



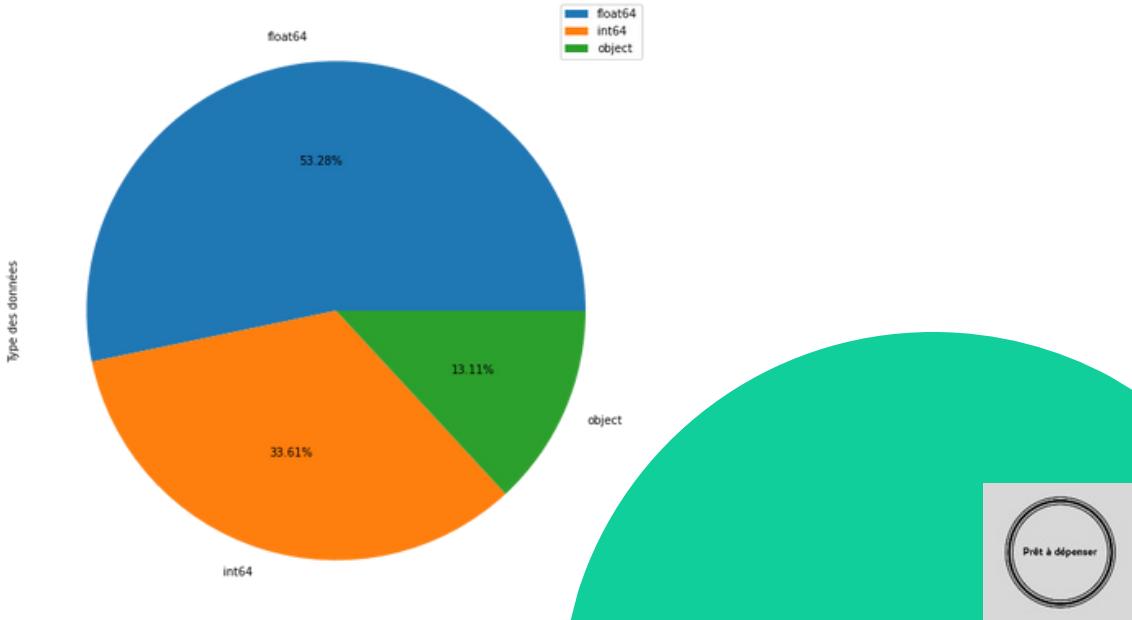
Analyse de la forme

Répartitions
des valeurs
vides

Répartition
des types
de données

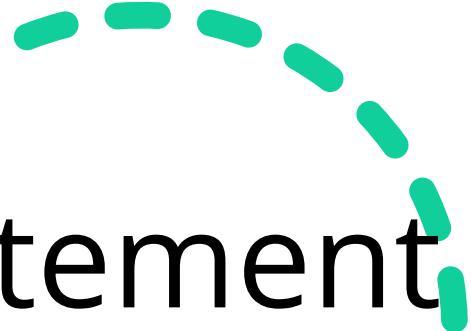


Répartition des types dans le jeu de données.

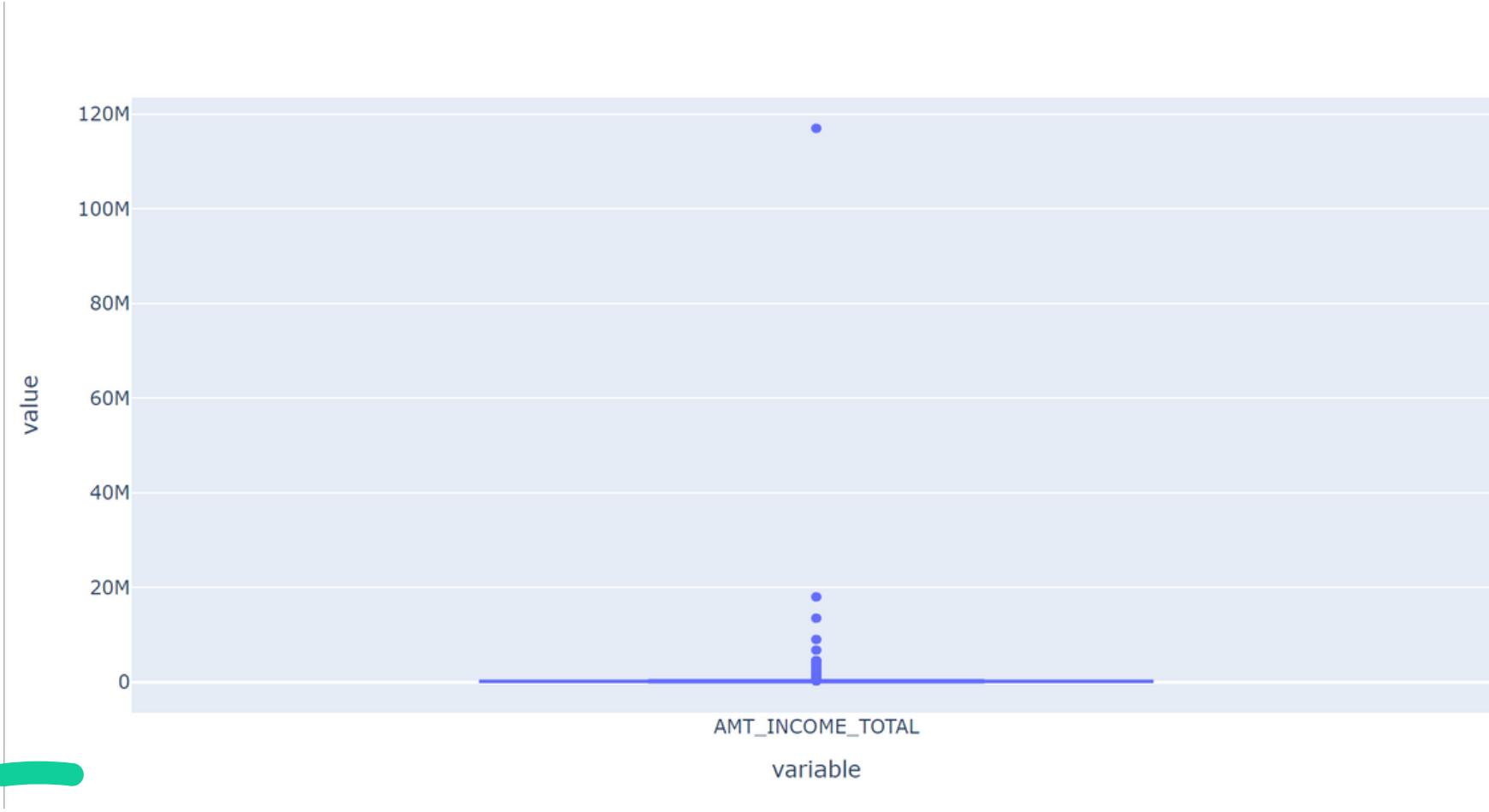


Prêt à dépenser

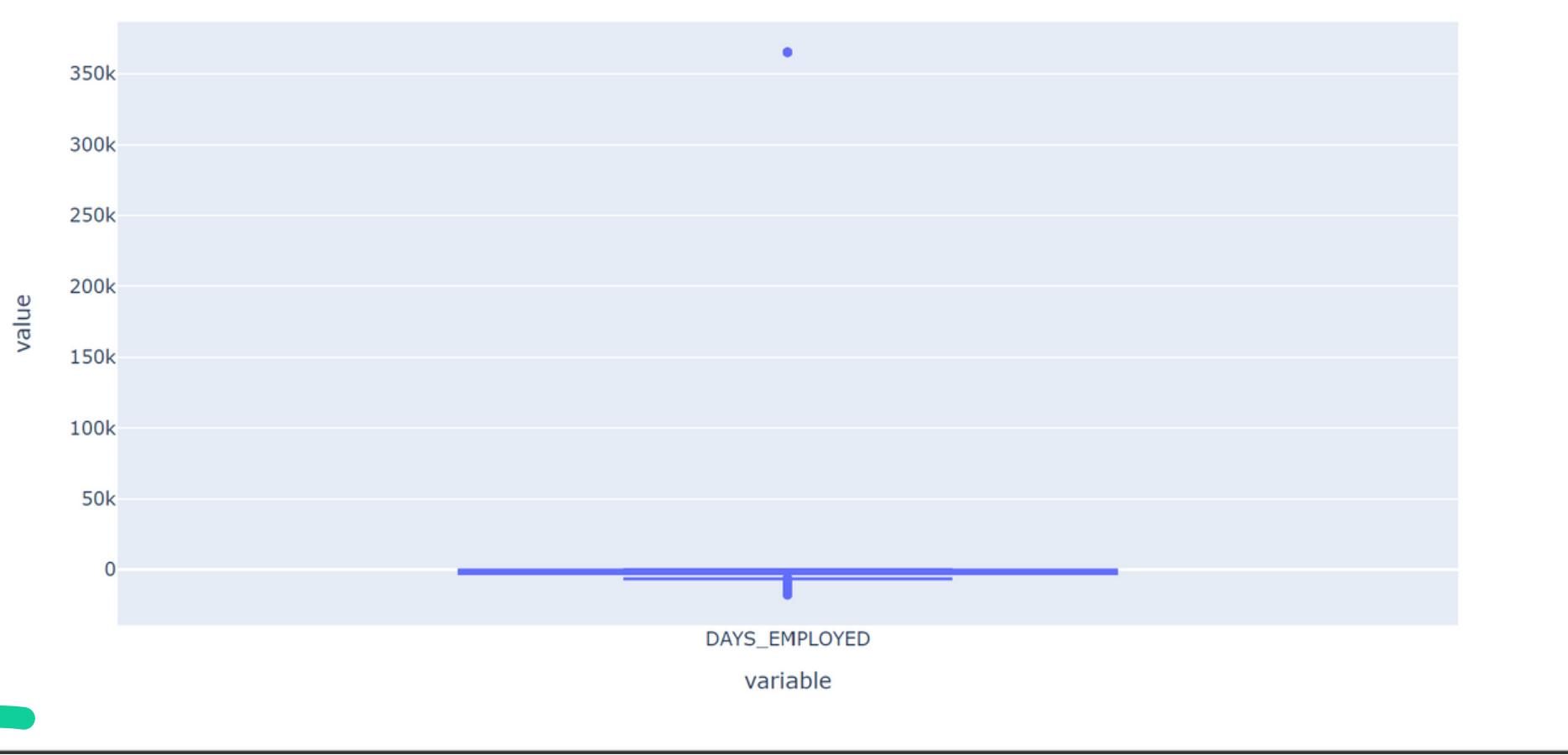
Traitement du jeu de données



Valeurs aberrantes

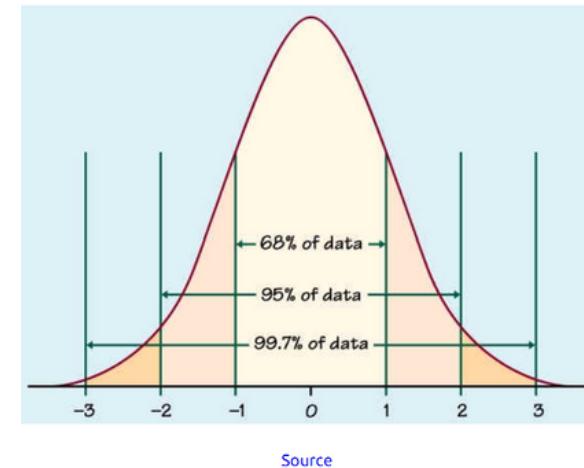
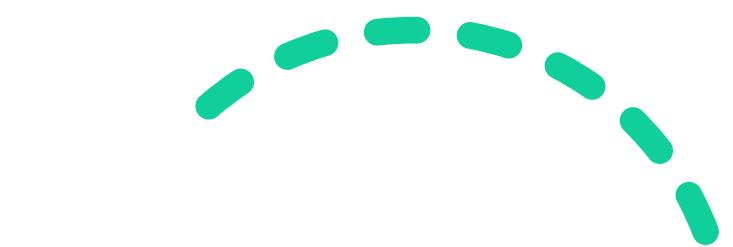


Valeurs aberrantes



Imputation

- Zscore (threshold 3)
- Days_col < Day_Birth
- Day_col > 0 = 0
- num_col.filna(0)
- cat_col.filna('Not Specified')



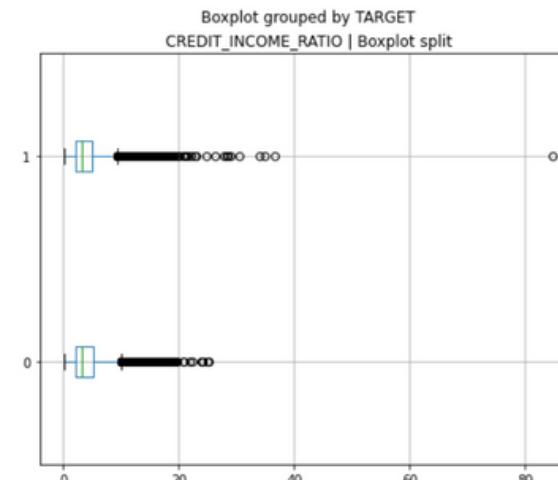
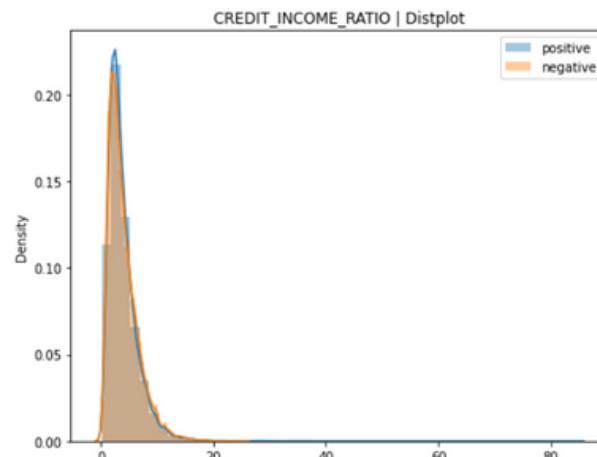
Standardisation et encodage des valeurs

- StandardScaler - pour les données quantitative
- One Hot Encoder - pour les données qualitative

AMT_REQ_CREDIT_BUREAU_MON	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR	x0_Cash loans	x0_Revolving loans	x1_F	x1_M	x1_XNA
-0.269936	-0.396749	-0.346733	1.0	0.0	0.0	1.0	0.0
-0.269936	-0.396749	-0.885567	1.0	0.0	1.0	0.0	0.0
-0.269936	-0.396749	-0.885567	0.0	1.0	0.0	1.0	0.0
-0.269936	-0.396749	-0.885567	1.0	0.0	1.0	0.0	0.0
-0.269936	-0.396749	-0.885567	1.0	0.0	0.0	1.0	0.0

Création de nouvelles colonnes

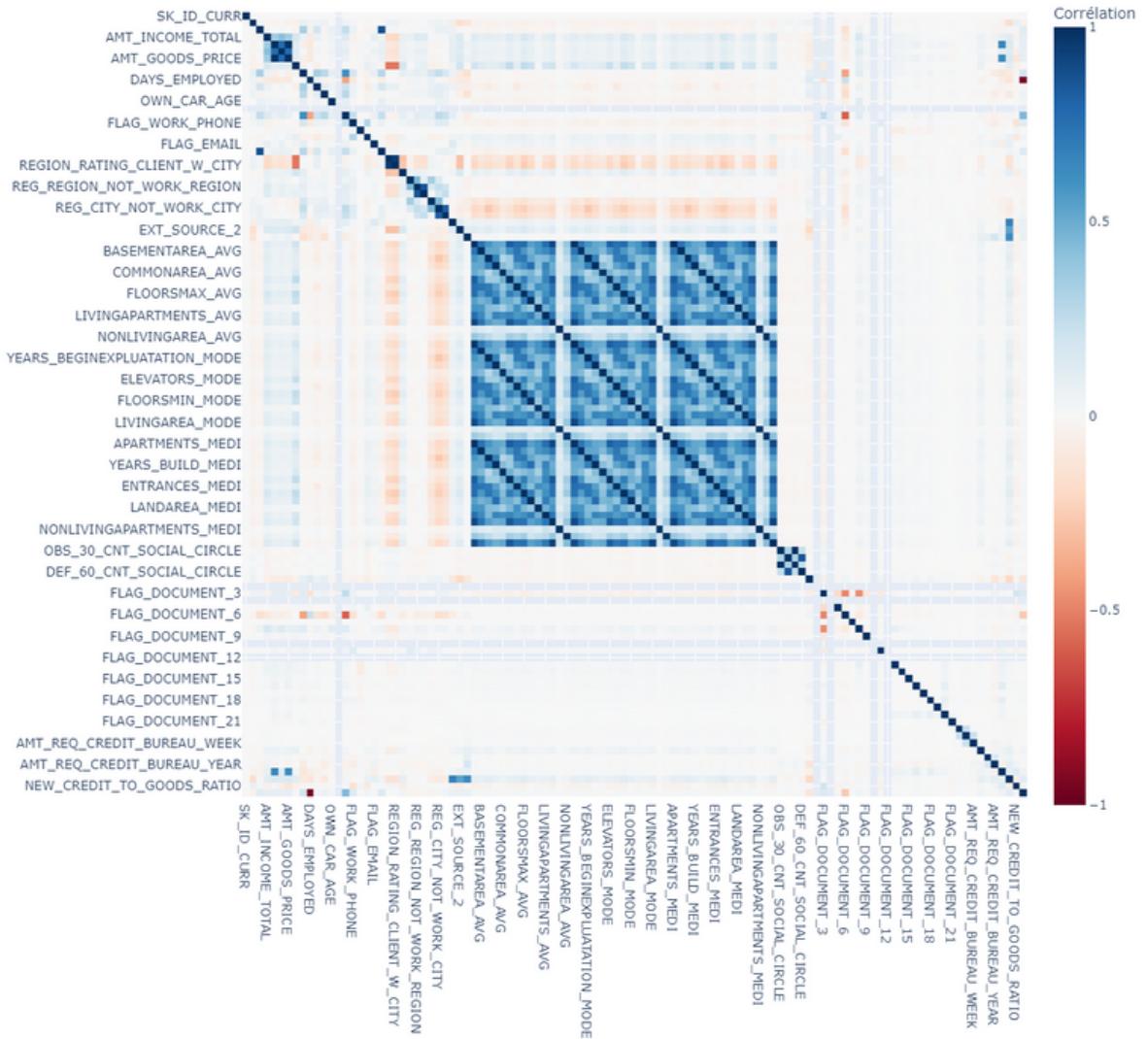
- CREDIT_INCOME_RATIO = AMT_CREDIT / AMT_INCOME_TOTAL
- CREDIT_GOODS_RATIO = AMT_CREDIT / AMT_GOODS_PRICE
- FLOORS_AVG = (FLOORSMAX_AVG + FLOORSMIN_AVG)/2



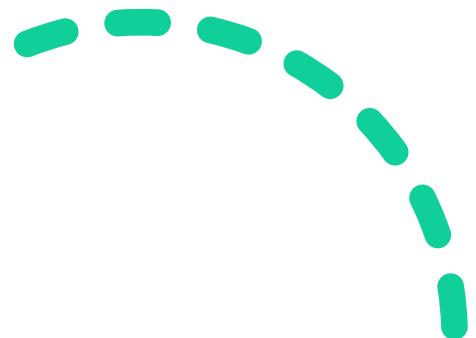
Corrélations entre les variables

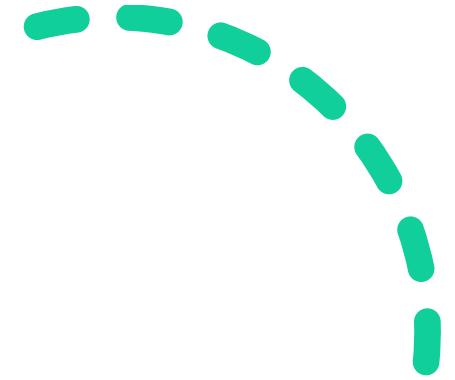
- Corrélations NaN: sont binaires
- Suprimé fortement corrélé (70%)

Corrélation entre les variables



Modélisation

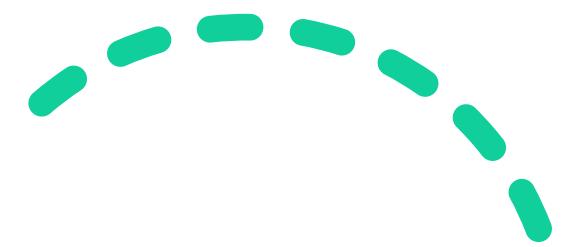




Métrique



- Recall
- Precision
- $F\beta$ -score (β à 2)
pour éviter d'accepter des clients qui ne rembourseront pas le prêt
=> Perte d'argent pour l'entreprise



Choix des modèles

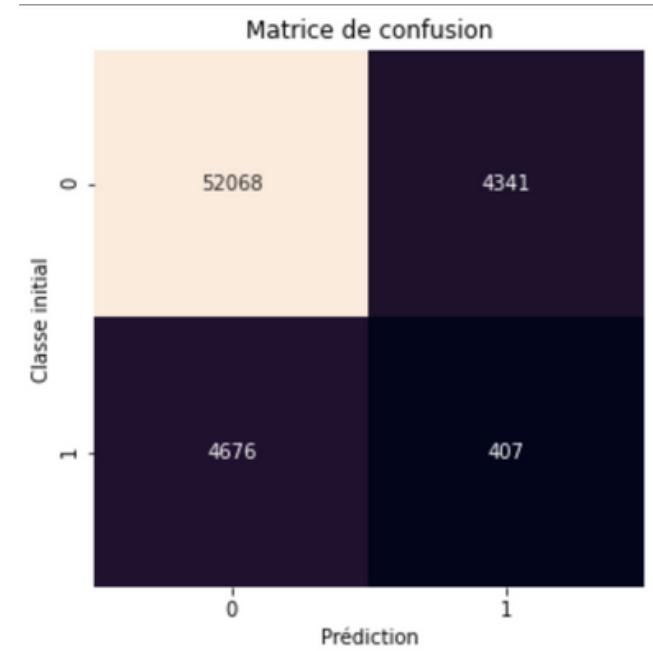
- Dummy Classifier
- Decision Tree Classifier
- Logistic Regression
- Random Forest

Confusion Matrix

Dummy Classifier

accuracy F2-score

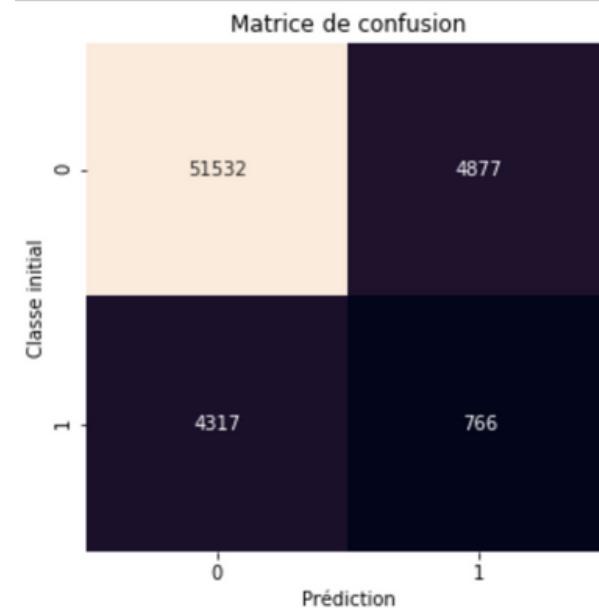
85.21% 8.36%



Decision Tree

accuracy F2-score

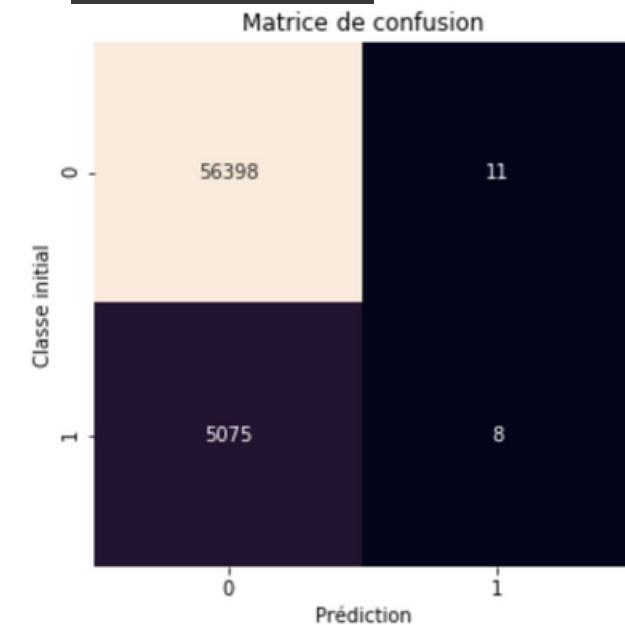
85.41% 15.31%



Logistic Regression

accuracy F2-score

91.97% 0.23%



Random Forest

accuracy F2-score

71.83% 39.41%

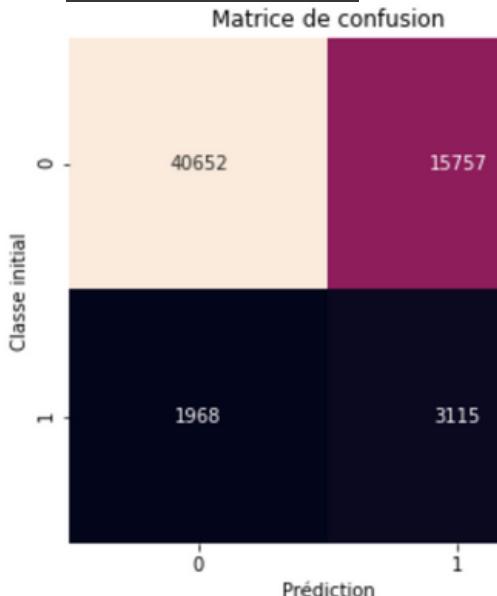
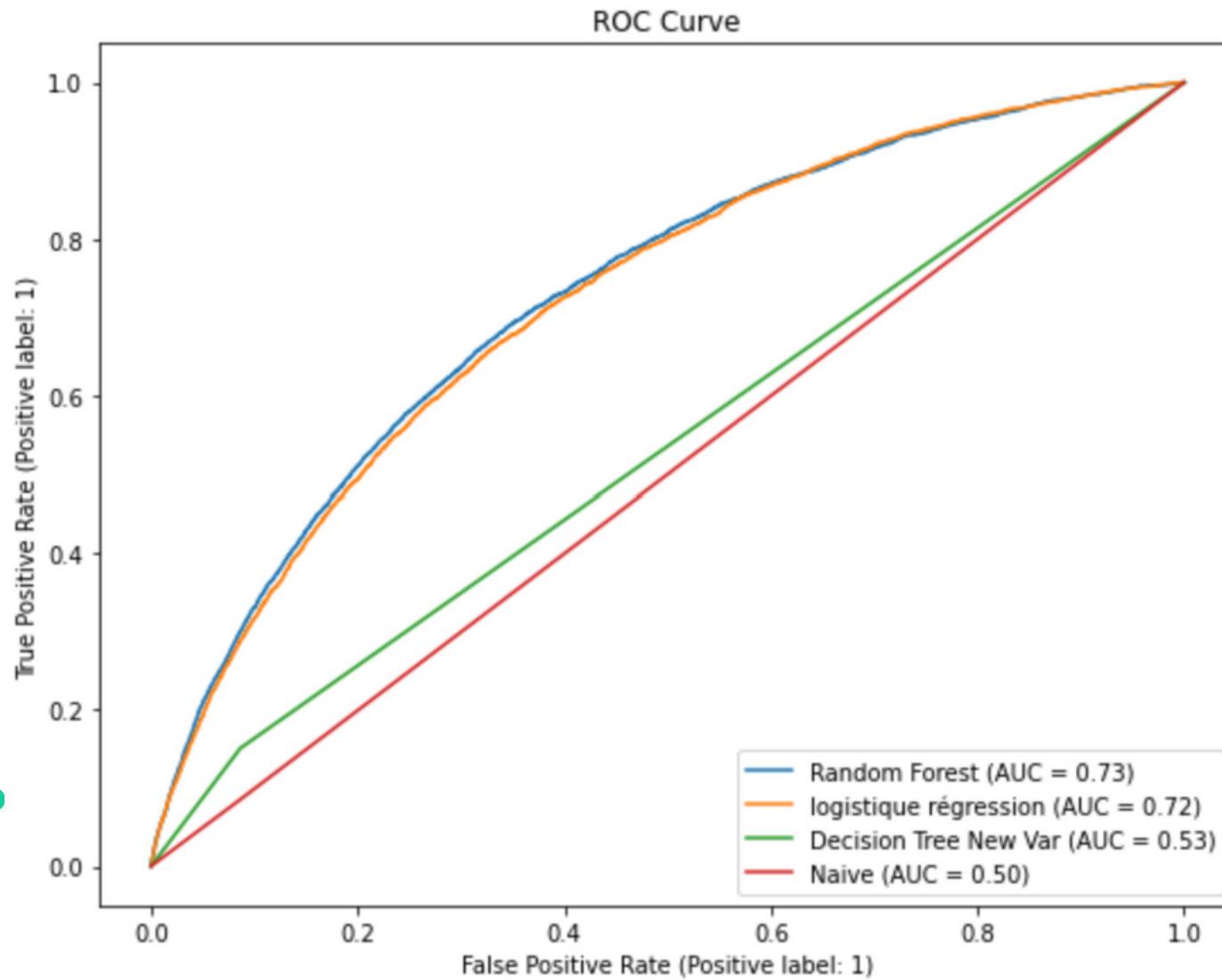


Tableau récapitulatif

	accuracy	F2-score
dummy	85.21%	8.36%
dtc basique	85.41%	15.31%
dtc nouvelles variables	85.36%	15.16%
régression logistique	91.97%	0.23%
Random Forest	71.83%	39.41%

ROC Curve



True Positive Rate (TPR) |

$$TPR = \frac{TP}{TP + FN}$$

False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN}$$

Hyperparamètres finaux

LogisticRegression



- logisticregression_C: 0.1
- logisticregression_max_iter: 1000
- logisticregression_solver: saga

Random Forest

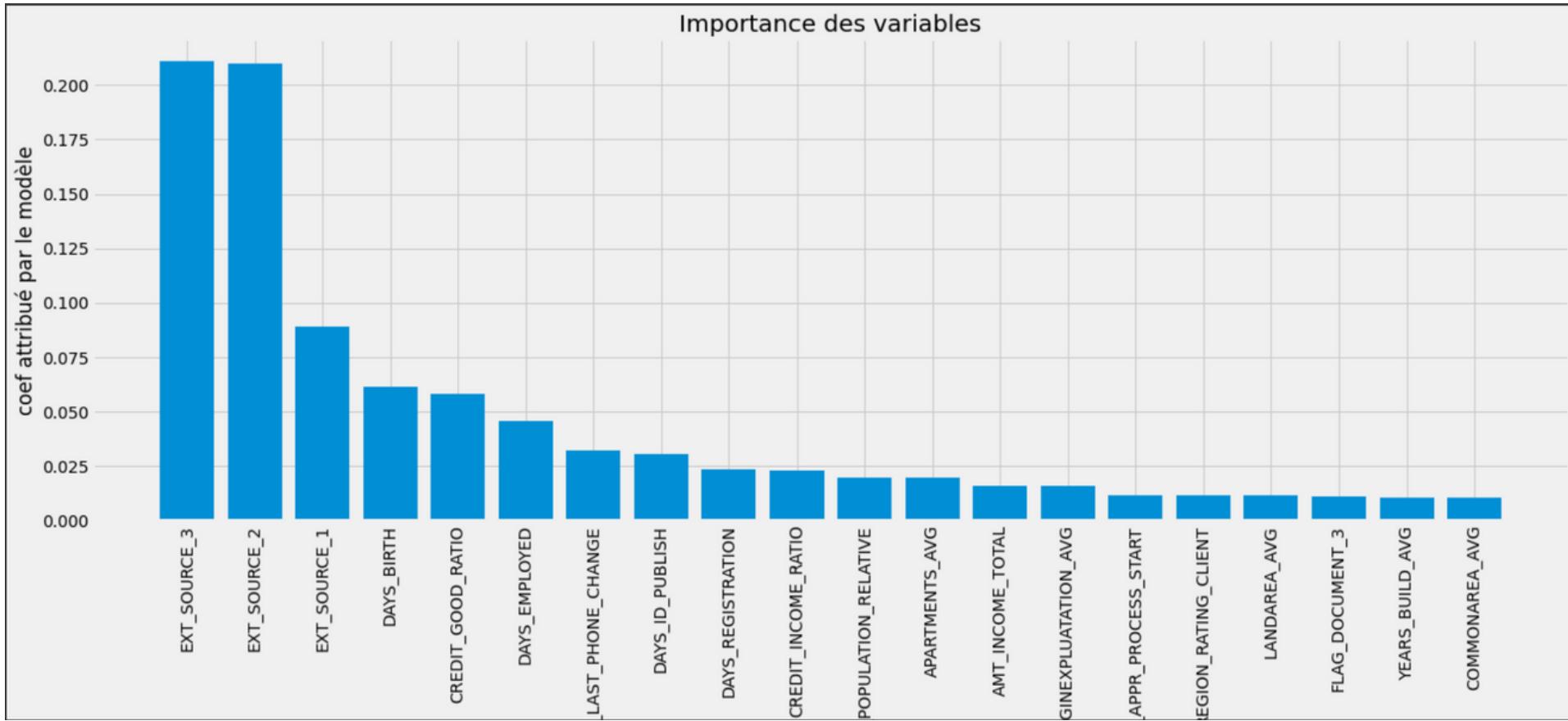


- randomforestclassifier_n_estimators: 250
- randomforestclassifier_class_weight: 'balanced'
- randomforestclassifier_criterion: 'entropy'
- randomforestclassifier_max_depth: 10
- randomforestclassifier_n_jobs: 60
- randomforestclassifier_random_state: 0

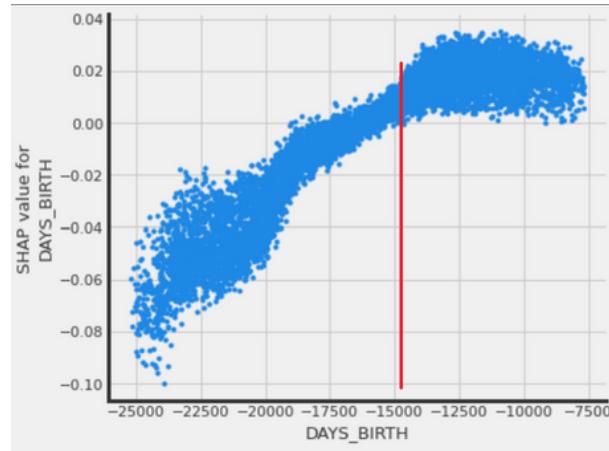
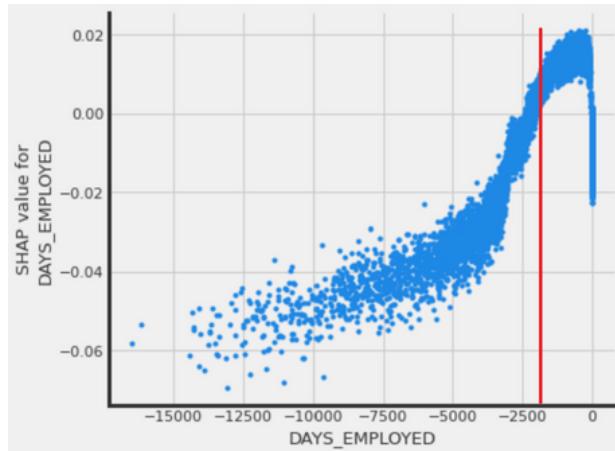
Explication de Model



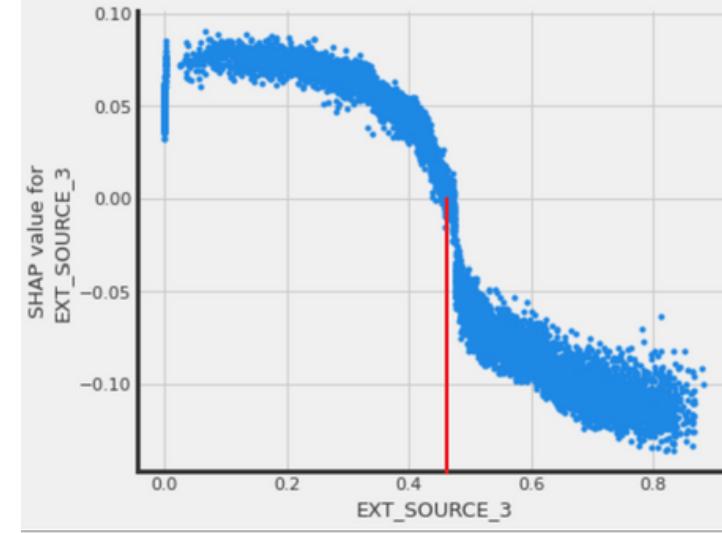
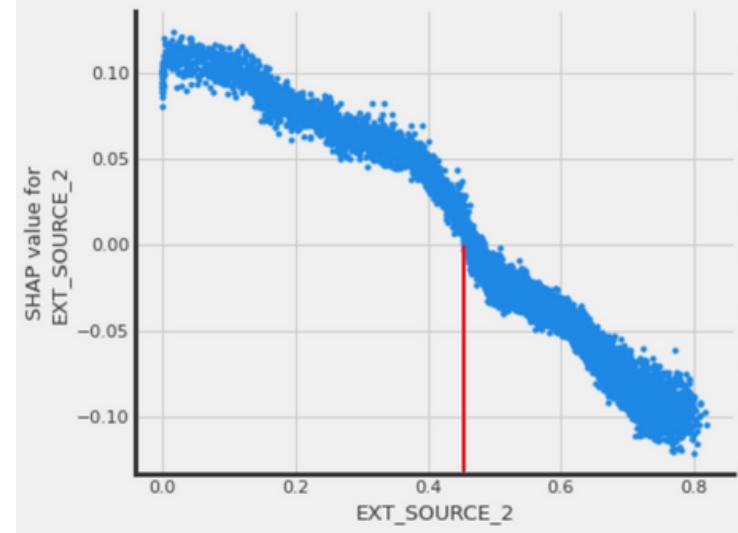
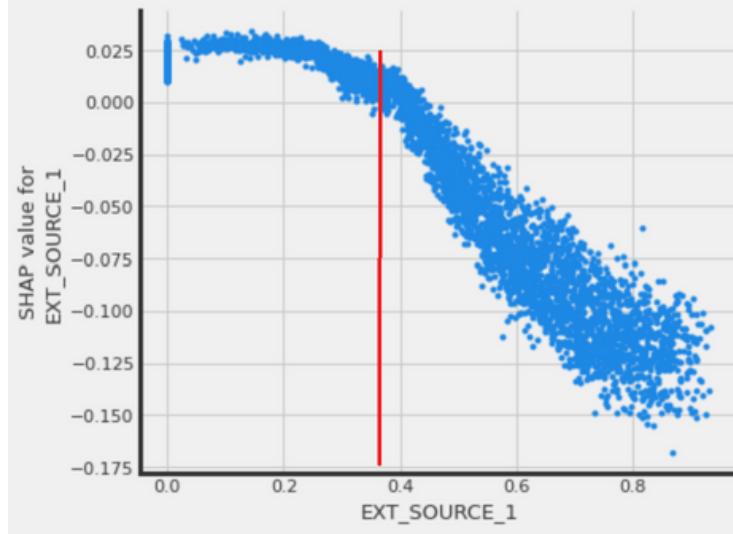
Feature Importance



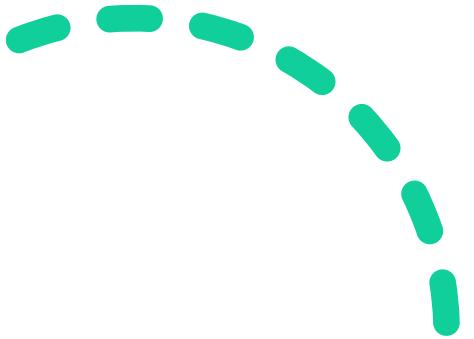
Explication SHAP



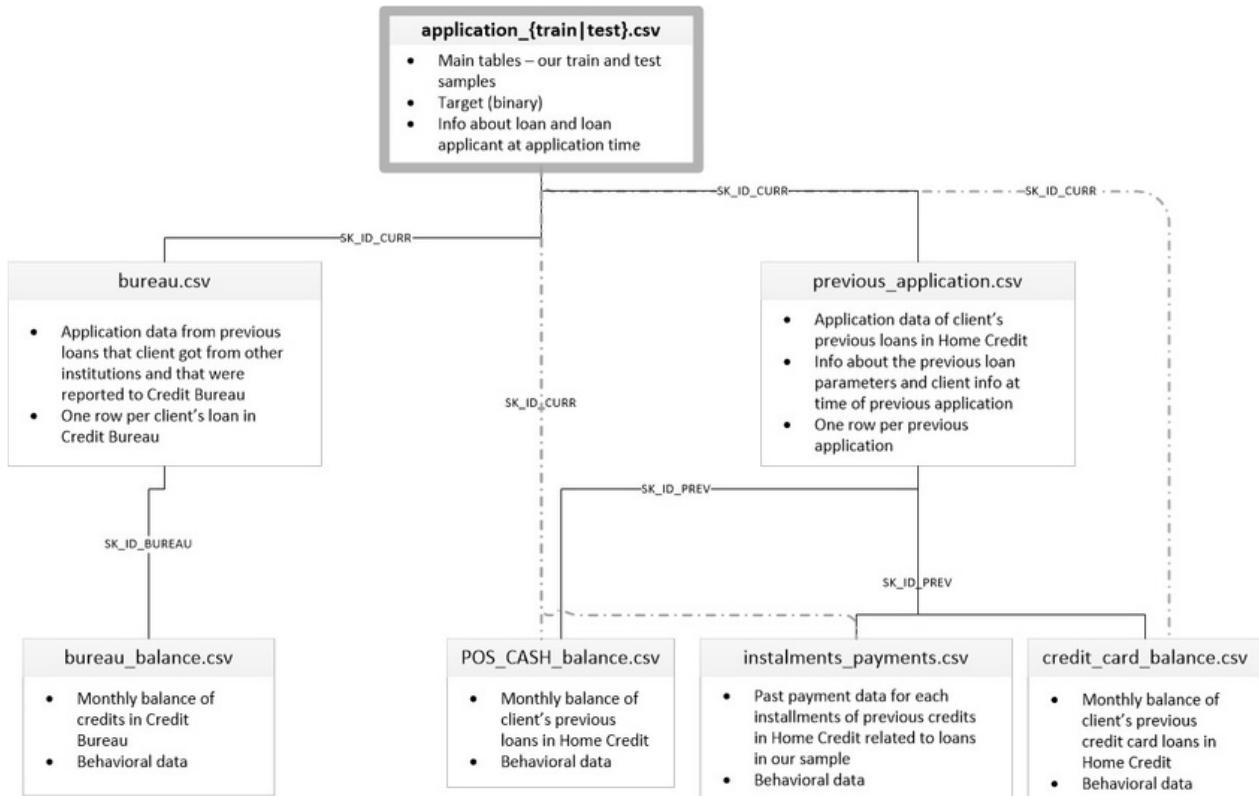
Explication SHAP



Conclusion



Conclusion



Accuracy: 71,83%
F2Score: 39.41%



Prendre tous les fichiers



Premier projet ML





Merci pour votre temps
et
votre attention.

Des questions ?

