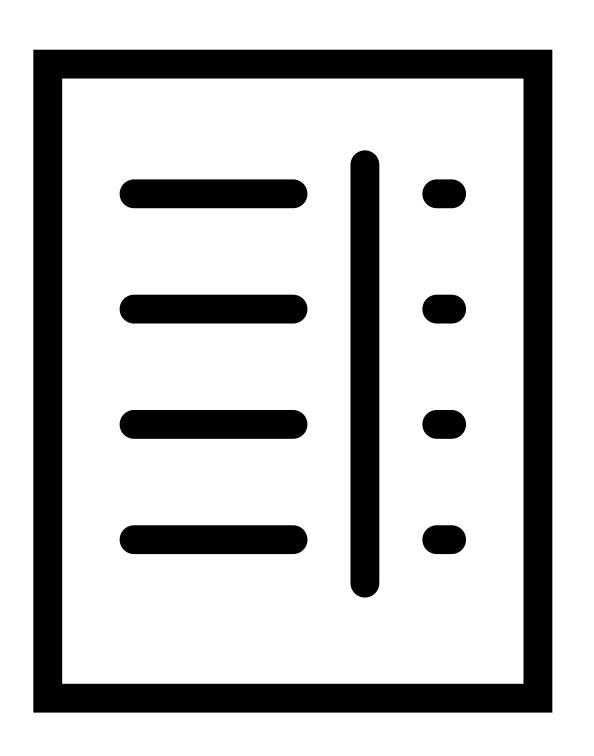




LAPRÉDICTION DES FAILLITES D'ENTREPRISES

Sommaire

- 1 Contexte
- 2 EDA
- 3 Preprocessing
- 4 Modèles
- 6 Conclusion





Présentation du Projet

Développeur dans une entreprise qui travaille dans le domaine de la FinTech.

Notre client est une société d'investissement qui souhaite détecter les faillites d'entreprises, ce qui leur permet de mieux comprendre et gérer les risques.

Nous avons créer un modèle de Machine Learning pour détecter la probabilité de faire faillite en utilisant plusieurs méthodes





Présentation du Dataset

Lignes

6819

Colonnes

96

Doublons

0

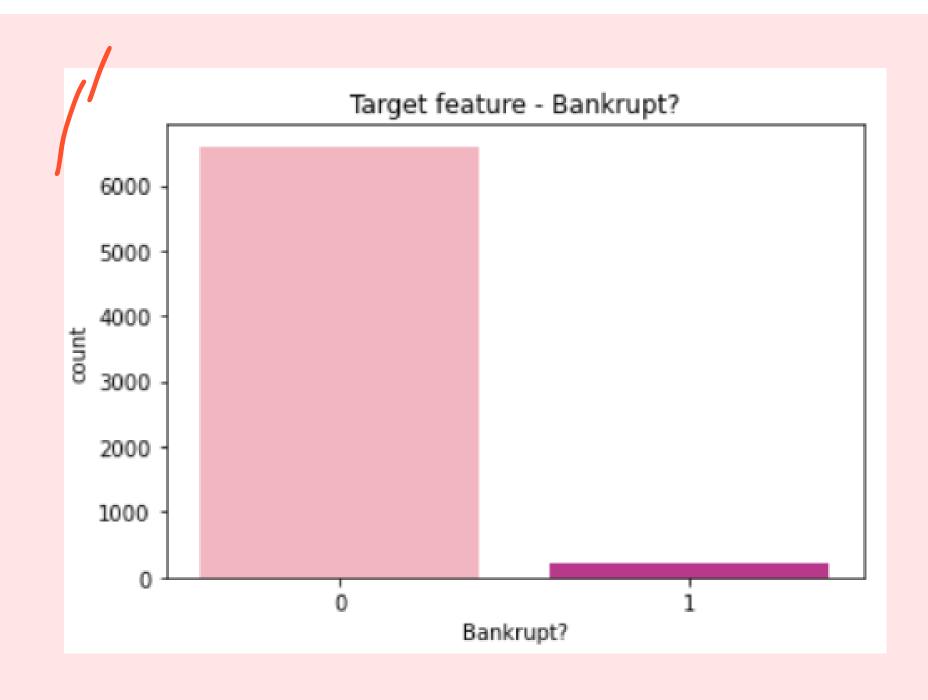
Valeurs manquantes

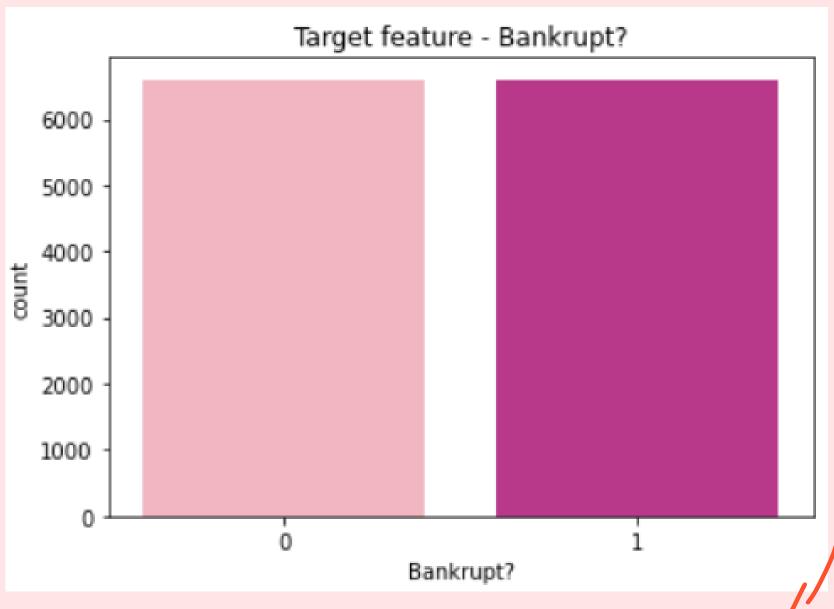
0

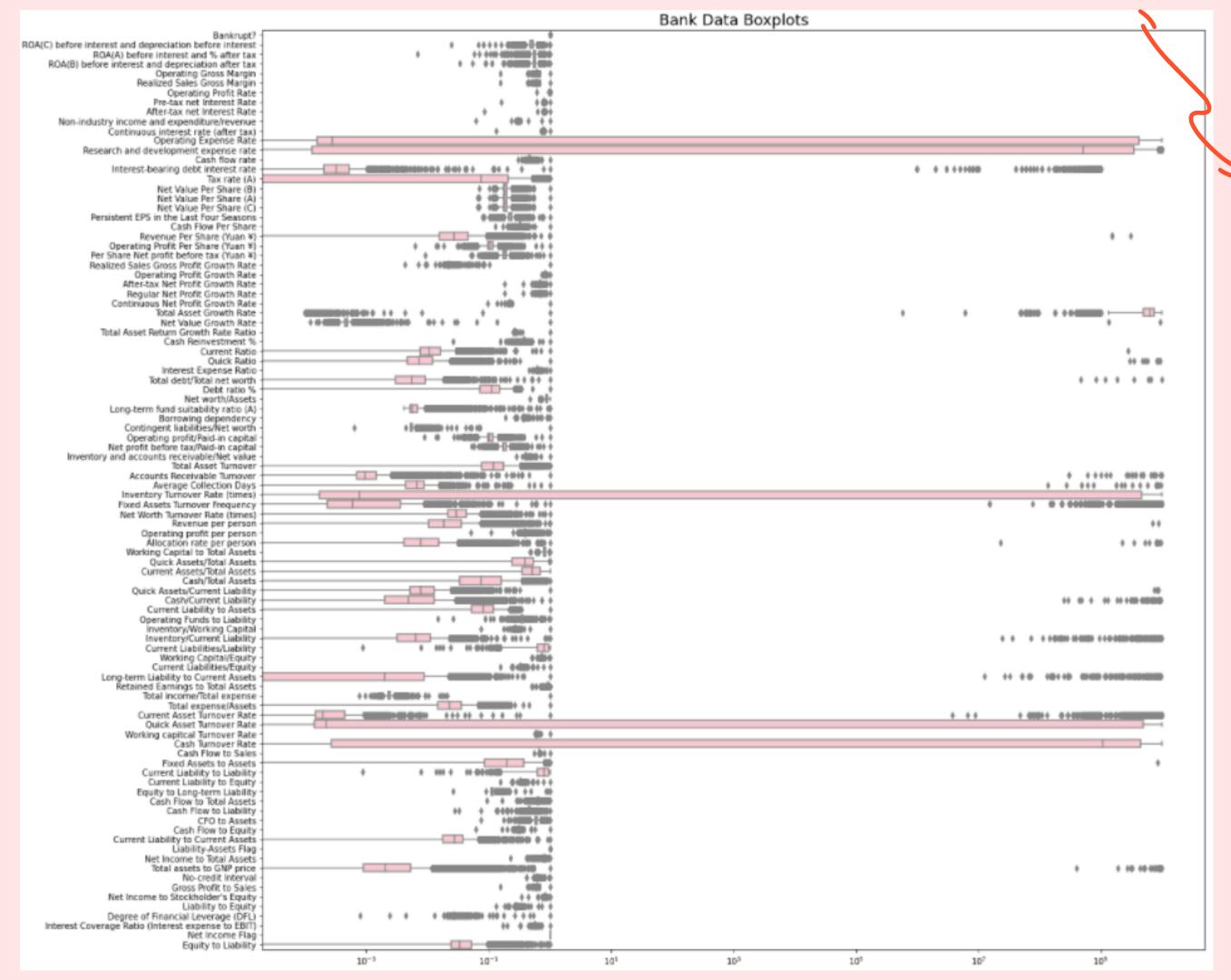




Sur-échantillonnage de la target à l'aide de SMOTE







Outliers



Forte présence d'outlier parmis la plupart des features

After-tax not interest Rate: Non-industry income and expenditure/revenue Continuous interest rate (after tax) Operating Expense Rate Research and development expense rate Cash flow rule: 1 Interest-bearing debt interest rate 1 Tox rate (A) 1 Net Volue Per Share (S) Net Value Per Share 16 Net Value Per Share (C) Persistent EPS in the Last Four Seasons Cash Flow Rer Share Revenue Per Share (Year V) Operating Profit Per Share (Nam V) Per Share Not profit before tax (Nam V) Realized Sales Gross Profit Growth Rate: Operating Frott Growth Rate After tax Net Profit Growth Rate Recular Net Profit Growth Rate Continuous Net Profit Growth Rate Total Asset Growth Rate -Not Value Growth Rate 4 Total Asset Return Growth Rate Ratio 4 Cosh Retrinolment % Coment Ratio Quick Ratio Interest Expense Ratio Debt ratio % -Not north/Reacts = Long-term fund safatistic rate (A) = Surrowing dependency Contingent habitities/flict worth Operating profit/Paid in-capital Inventory and accounts receivable/liet value Tital Asset Turnover . Accounts Receivable famover *** Avarage Collection Days * Invariary lamorer flate (fames): funed Asocia famore: Preparity Revenue per person Operating profit per person Allocation rate per person all Working Capital to Intal Associa Clark Associat Intal Associa Corrent Associa/Notal Associa Goldfeld Asoth Quick Assets/Current Hisbilley Carb/Current Hability Current Liability to Assets Operating Busde to Hability 40 Inventory/Merking Capital 4 Invertory/Carrent Leability 4 Current Lubritins/Eurotity Current Lisbillries/Coulty Long-term Liability to Current Assets | | | | Total income/Total expense Mal operacitacs 4 Carrent Asset, lumover flate+ Quick Asset Turnover Rate Working capitcal lumover flate -Carth Tumover Rate Road Assets to Assets Current Liability to Liability Carrent Liability to Equity #1 Cash Flow to Total Assets * Cash Flow to Liability Carth Firm to Equity Current Liability to Current Ansets Net Income to Total Assets

ROA(C) before interest and depreciation before interest

ROA(A) before interest and % after tax ROA(B) before interest and depreciation after tax

Operating Gross Margin -Realized Sales Gross Margin Operating Profit Ratu Pro-tax not Interest Ratu

Cross Profit to Sales

Degree of Financial Leverage (DR.)

Net Income to Stockholder's Equity

Interest Coverage Ratio (Interest expense to ESIT)

1.00

- 0.75

0.50

+ 0.25

0.00

0.25

-0.50

0.75

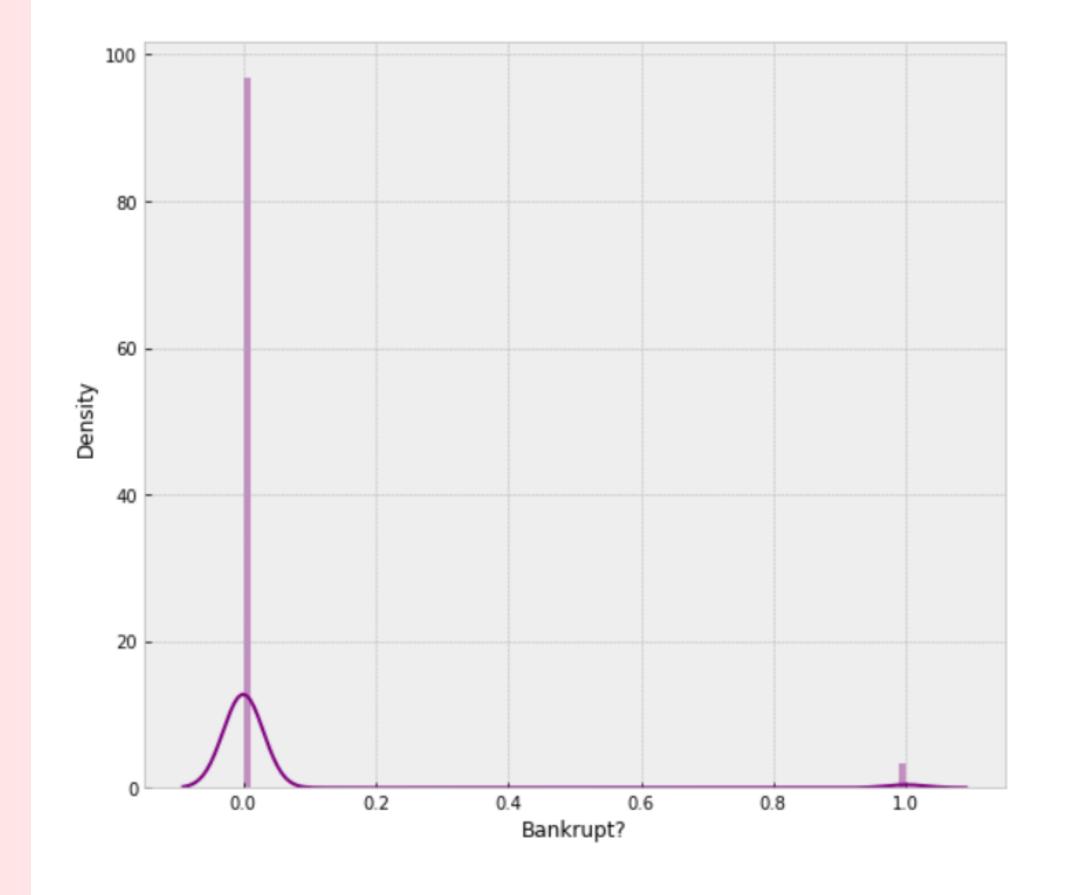


Analyse de la target

Notre cible est énormément déséquilibrée

Seulement 3% de 1 (faillites)

Contre 97% de 0 (pas de faillites)

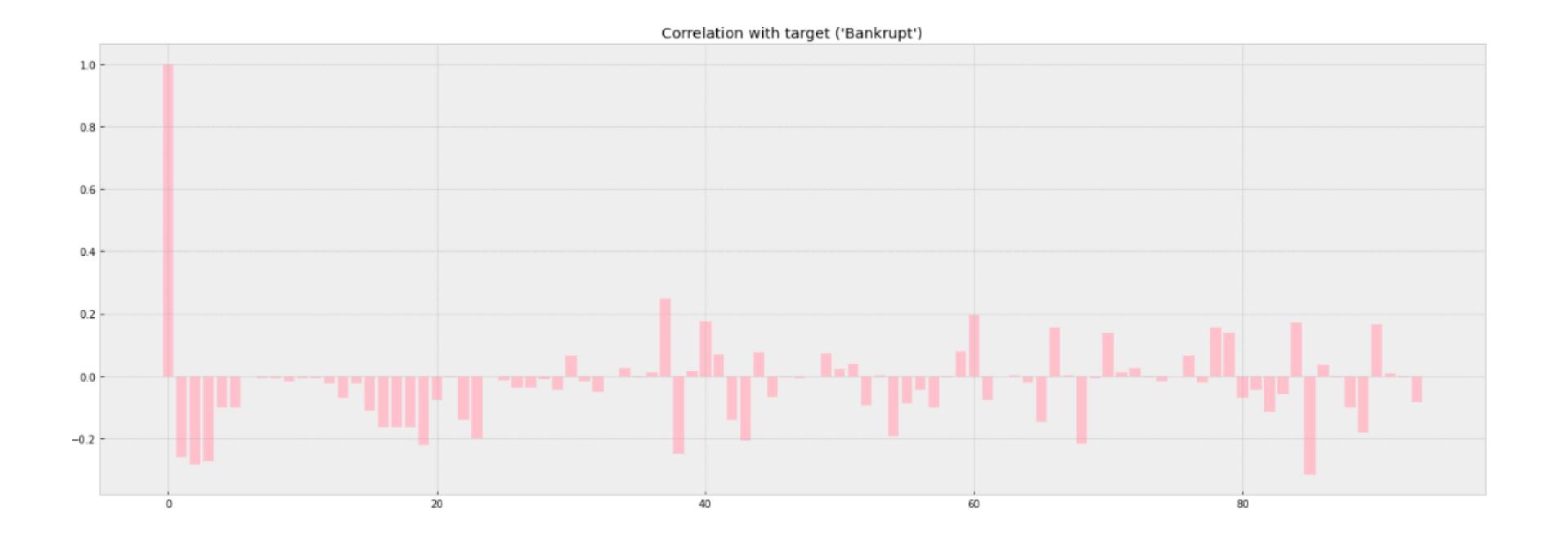




Corrélation avec la target



Plusieurs features sont corrélées positivement et négativement avec notre cible.



PREPROCESSING



Features Selection

Il y a 22 variables corrélés à 0.15% avec la cible:	
Bankrupt?	1.000000
Debt ratio %	0.250161
Current Liability to Assets	0.194494
Borrowing dependency	0.176543
Current Liability to Current Assets	0.171306
Liability to Equity	0.166812
Current Liabilities/Equity	0.153828
Current Liability to Equity	0.153828
Net Value Per Share (C)	-0.164784
Net Value Per Share (B)	-0.165399
Net Value Per Share (A)	-0.165465
Net Income to Stockholder's Equity	-0.180987
Working Capital to Total Assets	-0.193083
Per Share Net profit before tax (Yuan ¥)	-0.201395
Net profit before tax/Paid-in capital	-0.207857
Retained Earnings to Total Assets	-0.217779
Persistent EPS in the Last Four Seasons	-0.219560
Net worth/Assets	-0.250161
ROA(C) before interest and depreciation before interest	-0.260807
ROA(B) before interest and depreciation after tax	-0.273051
ROA(A) before interest and % after tax	-0.282941
Net Income to Total Assets	-0.315457
Name: Bankrupt?, dtype: float64	

Préprocessing et Techniques



Split des données avec StratifiedKFold



Scaling avec Robust Scaler



Balancement sur le X-train et Y-train avec la méthode de SMOTE

KODELES



Itération des différents modèles



```
Fitting 5 folds for each of 18 candidates, totalling 90 fits
0.9030497611157557
{'model_learning_rate': 1, 'model_n_estimators': 70}
                        recall f1-score support
             precision
                  0.99
                           0.93
                                     0.96
                                              1319
                  0.26
                          0.73
                                     0.39
                                                44
                                     0.93
                                              1363
   accuracy
  macro avg
                  0.63
                           0.83
                                     0.67
                                              1363
weighted avg
                  0.97
                           0.93
                                              1363
                                     0.94
[[1229
        90]
                             Gradient Boosting
   12
        32]]
```

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
0.9222927851250491
{'model n neighbors': 2, 'model weights': 'uniform'}
                          recall f1-score support
              precision
                  0.98
                            0.95
                                      0.96
                                                1319
                                      0.23
                  0.18
                                                  44
                            0.32
                                      0.93
                                                1363
    accuracy
                            0.63
                                      0.60
                                                1363
   macro avg
                  0.58
weighted avg
                  0.95
                            0.93
                                      0.94
                                                1363
[[1254
         65]
                                                 KNN
         14]]
    30
```

```
Fitting 5 folds for each of 8 candidates, totalling 40 fits
0.9039671907682907
{'model max depth': None, _'model__splitter': 'random'}
                         recall f1-score
             precision
                                            support
                  0.98
                            0.92
                                     0.95
                                               1319
                                     0.24
                  0.16
                           0.45
                                                 44
                                     0.91
                                               1363
   accuracy
  macro avg
                  0.57
                            0.69
                                     0.60
                                               1363
weighted avg
                  0.95
                                     0.93
                                               1363
                           0.91
[[1217 102]
                                 Decision Tree
       20]]
   24
```

```
Fitting 5 folds for each of 9 candidates, totalling 45 fits
0.9079968305449515
{'model class weight': 'balanced subsample', 'model max depth': 10, 'model n estimators': 200
             precision
                         recall f1-score support
                 0.99
                          0.94
                                    0.97
                                             1319
                                               44
          1
                  0.31
                          0.84
                                    0.45
                                    0.93
                                             1363
    accuracy
   macro avg
                           0.89
                                    0.71
                                             1363
                 0.65
                                    0.95
                                             1363
weighted avg
                 0.97
                           0.93
        82]
[[1237
                                                      Random Forest
    7 37]]
```

```
Fitting 5 folds for each of 1 candidates, totalling 5 fits
0.9202761186262561
{'model base estimator': RandomForestClassifier(class weight='balanced subsample', max depth=10,
                     n estimators=250), 'model n estimators': 150}
            precision recall f1-score support
                 0.99
                          0.95
                                   0.97
                                            1319
                 0.34
                          0.75
                                   0.47
                                              44
                                   0.95
                                            1363
   accuracy
  macro avg
                 0.67
                          0.85
                                   0.72
                                            1363
weighted avg
                 0.97
                          0.95
                                   0.96
                                            1363
[[1256
        63]
                                  AdaBoost + RandomForest
 [ 11 33]]
```

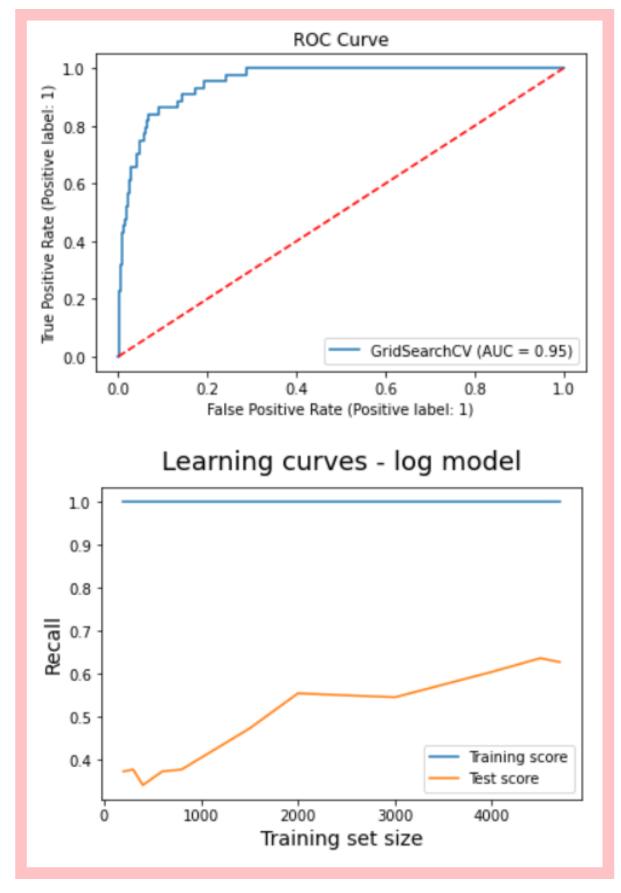
```
Fitting 5 folds for each of 3 candidates, totalling 15 fits
0.7397523191748674
 ['model__learning_rate': 'adaptive', 'model__learning_rate_init': 0.1, 'model__max_iter': 400}
             precision
                       recall f1-score support
          0
                 1.00
                          0.74
                                   0.85
                                             1319
                 0.10
                          0.91
                                   0.19
                                              44
                                   0.75
                                            1363
   accuracy
                                   0.52
   macro avg
                 0.55
                          0.82
                                            1363
                                   0.83
weighted avg
                 0.97
                          0.75
                                            1363
[[976 343]
                                  Multi Layer Perceptron
   4 40]]
```

```
Fitting 5 folds for each of 1 candidates, totalling 5 fits
0.9266904040047944
{'model final_estimator': LogisticRegression(class_weight='balanced')}
             precision
                         recall f1-score
                                             support
                            0.95
                   0.99
                                      0.97
                                                1319
                  0.34
          1
                                      0.48
                                                  44
                            0.82
                                      0.94
                                                1363
    accuracy
                            0.88
                                      0.72
                                                1363
                  0.67
   macro avg
weighted avg
                  0.97
                                      0.95
                                                1363
                            0.94
[[1249
        70]
                                                    Stacking
        36]]
     8
```

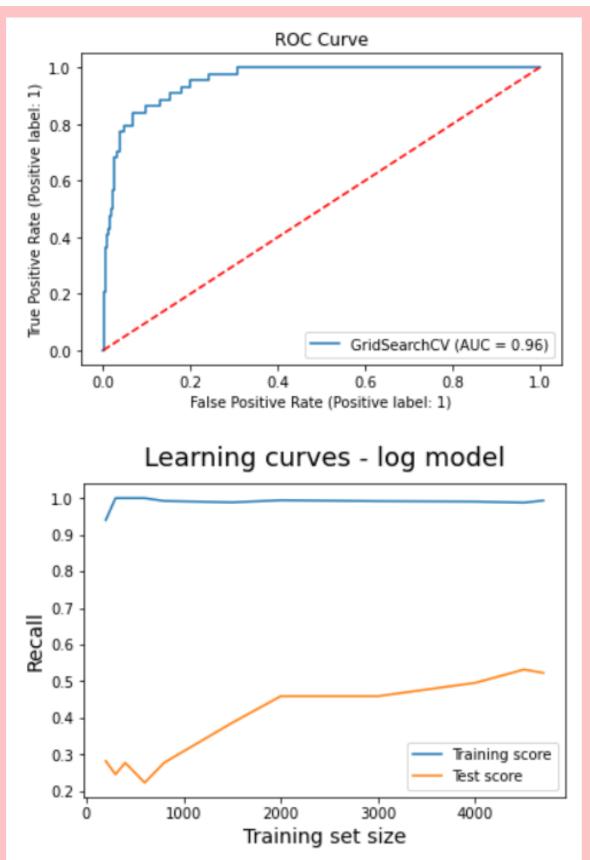


Evaluation des meilleurs modèles

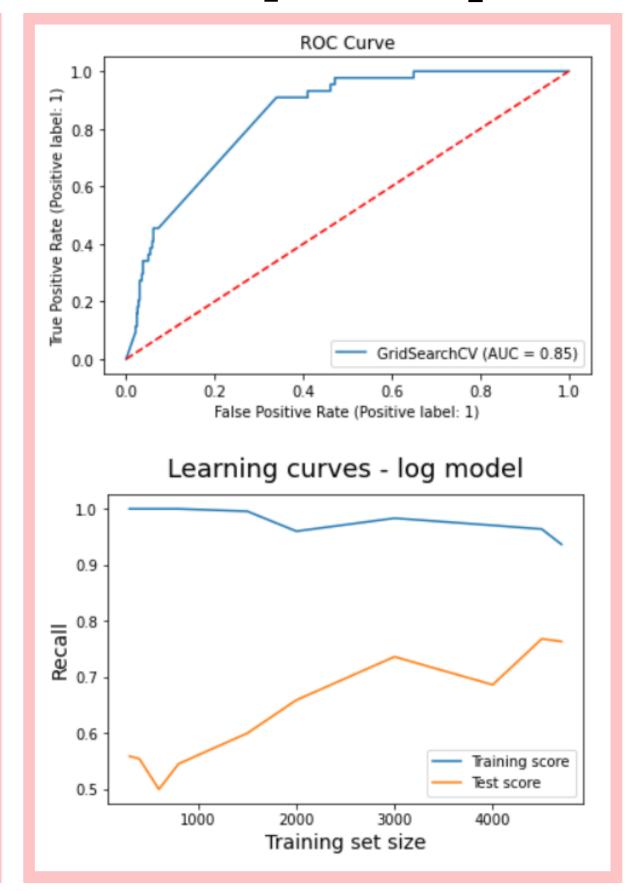
RandomForest



Stacking



Multi Layer Perceptron



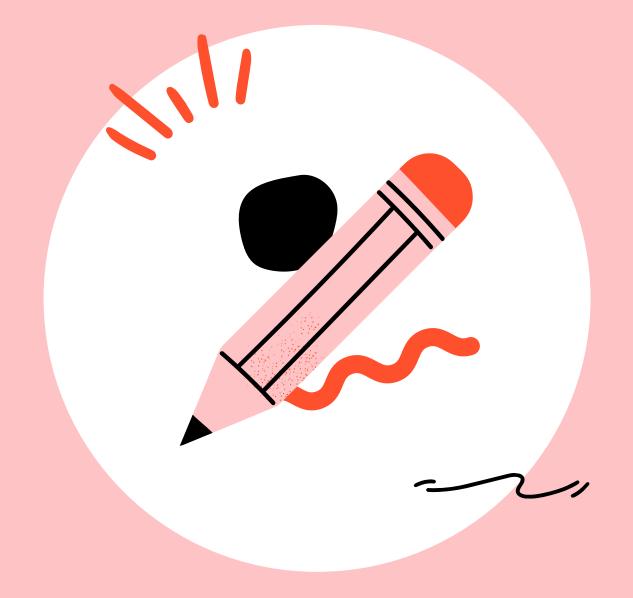
Conclusion



Le déséquilibre de la cible entraîne une mauvaise prédiction du model.



Grâce à certaines techniques de balancement il est possible de palier à ce problème.





La plupart des modèles nous permette d'obtenir une bonne reconnaissance de la classe en infériorité numérique.



Mais cela ce fait au détriment de la reconnaissance de la classe dominante.