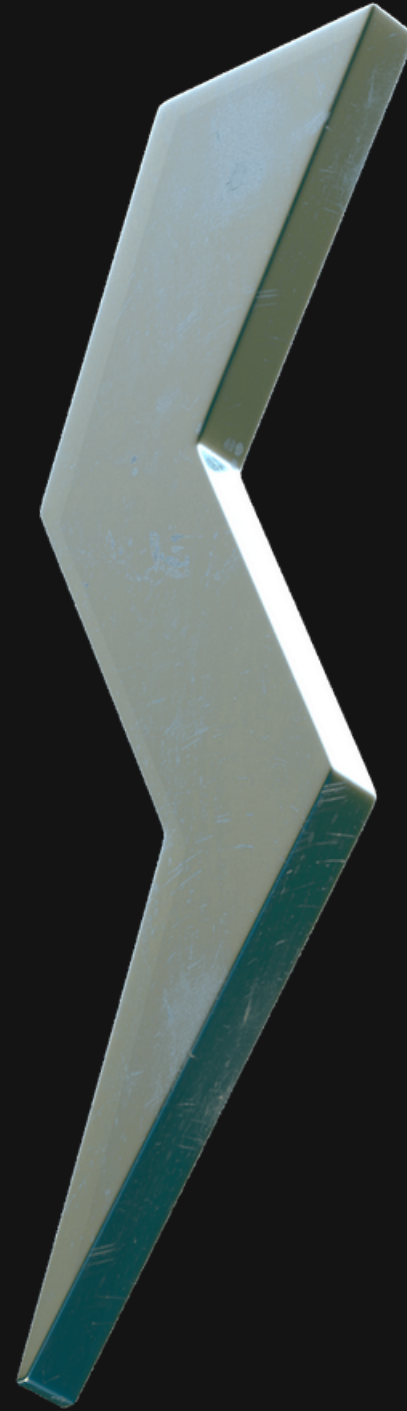


Loan Prediction

Will you get a loan ?

Présentation dataset



2 datasets j'ai utilisées le dataset
train
pour mes models et l'eda

pas de duplicates

160 valeurs nulle

beaucoup de outliers dans toutes les
colonnes

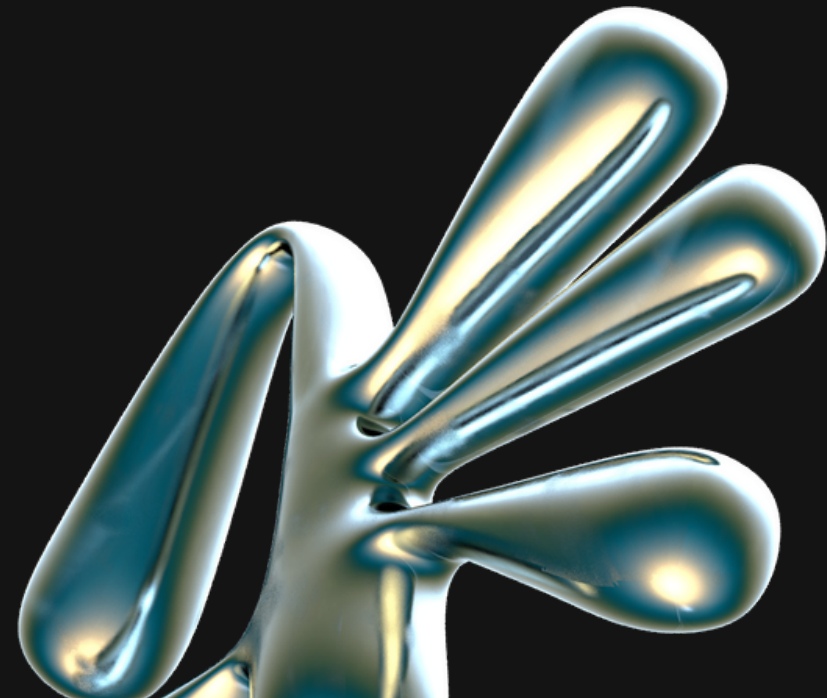
Colonnes 13
614 lignes

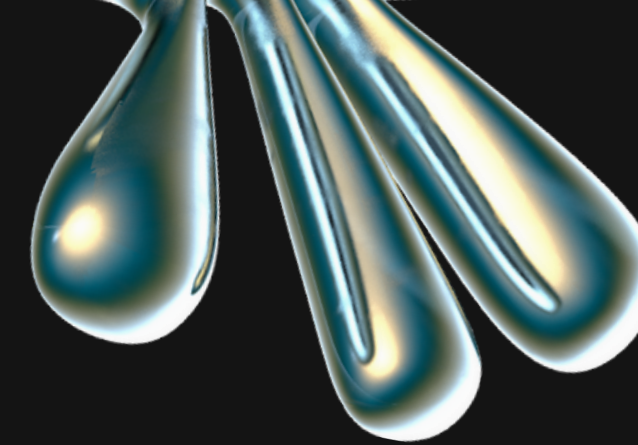
les différents model

- dummy classifier + standard scaler et minmax scaler
- Logistic regression + Smote
- Logistic regression + standard scaler et minmax scaler
- KNN + smote

Le premier résultat doit être petit tant dit que le second le plus proche de 1

- 0.54 et 0.51 très mauvais
- 0.46 et 0.76 (était à 0.80) meilleur modèle
- 0.39 et 1.00 overfitting
- 0.36 et 0.60 pas mauvais mais très gros taux d'erreur





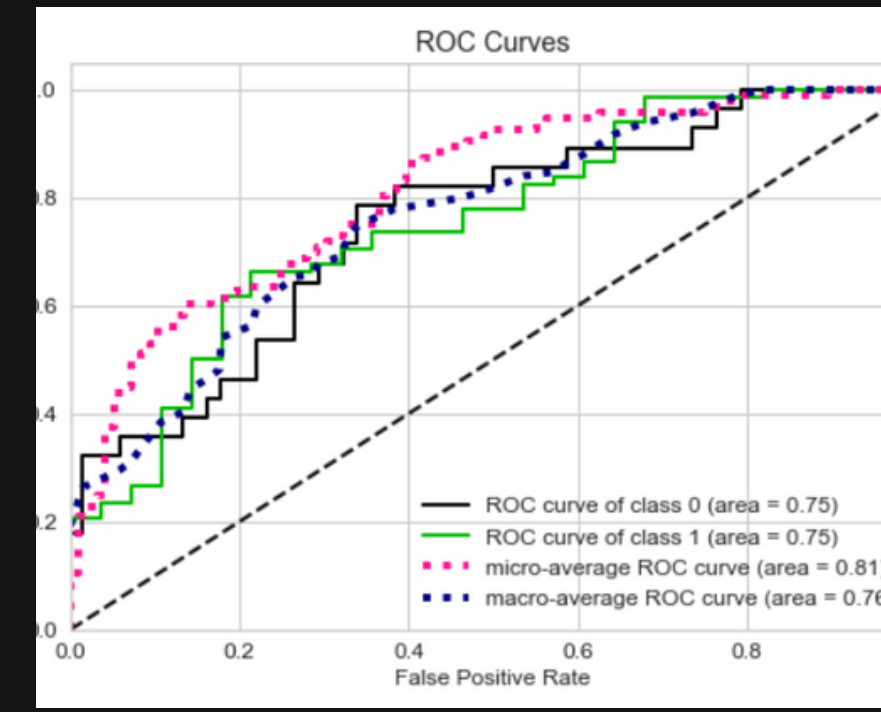
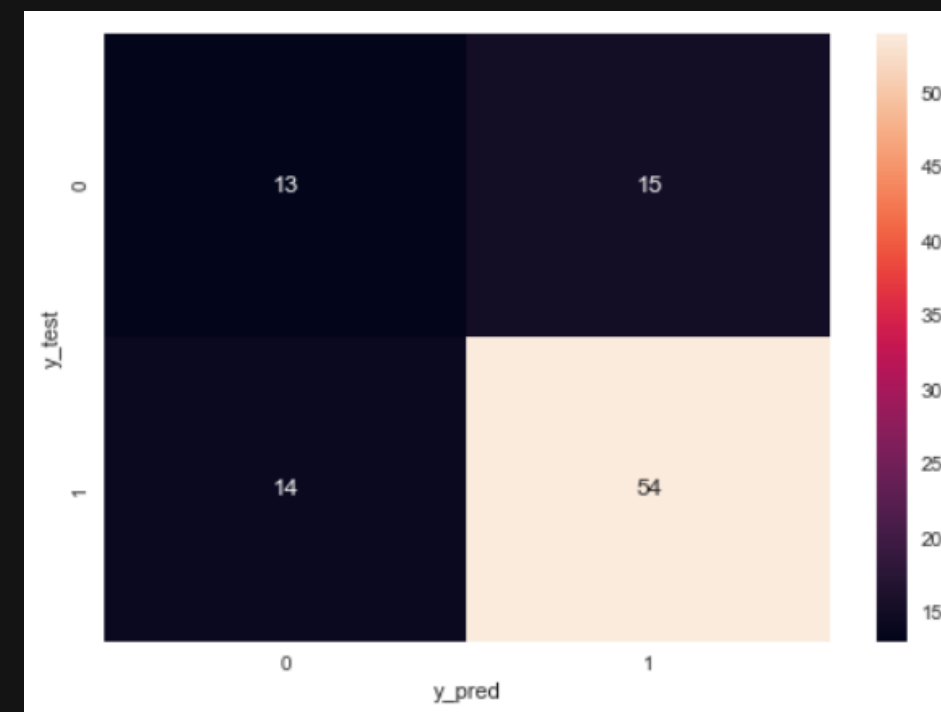
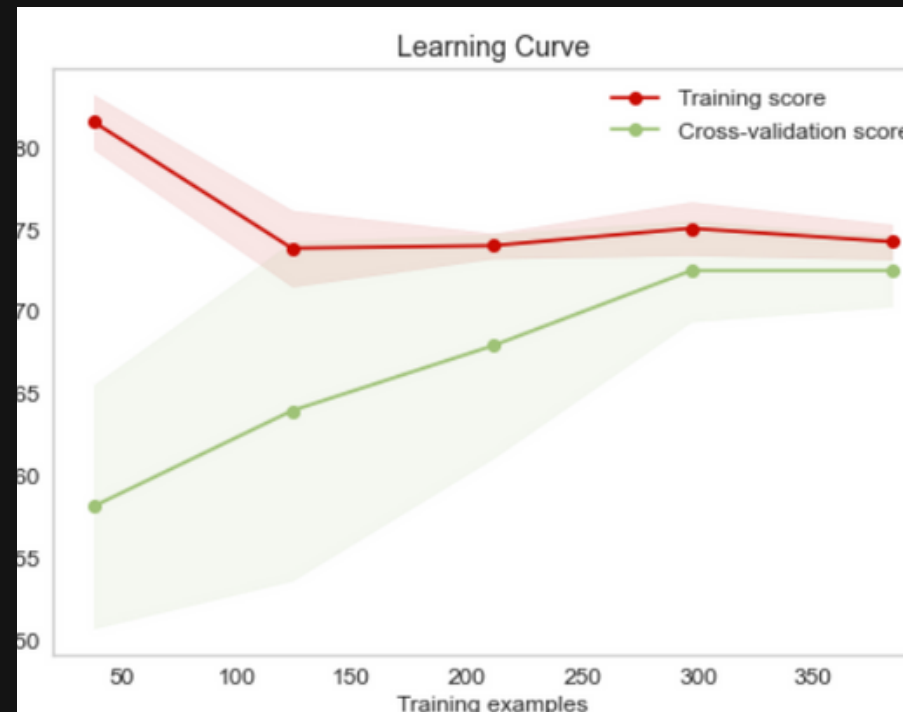
Best model

Logistic Regression + Smote

recall

0.46

0.79



Amélioration : plus de temps , plus de data ...

Conclusion

le type de personne susceptible
d'avoir un prêt

- pas marié
- dépendant
- job stable
- salaire entre 0 et 25000
- habite dans un endroit Semiurbain

