

Projet 8

**Réalisez un Dashboard
Et assurez une veille
technique**

Rédigé et présenté par

Etudiant : Gassuc Cédric

Mentor : Mohammed El Abrid

Introduction

Problématique : Comment répondre à la demande croissante de transparence des clients concernant les décisions d'octroi de crédit ?

Objectif : Fournir aux chargés de relation client un outil intuitif et interactif pour faciliter l'explication des décisions d'octroi de crédit, afin de renforcer la transparence et instaurer une relation de confiance avec les clients lors des rendez-vous.

Plan

1

Fonction de cout métier

2

Shap globale & Shap local

3

Dashboard

4

**Réalisez une veille technique :
Analyse des sentiments**

5

Modèle classique : TF-IDF + Naïve bayes

6

Modèle récent : DistilBert

6

Comparaison des modèles

6

Conclusion

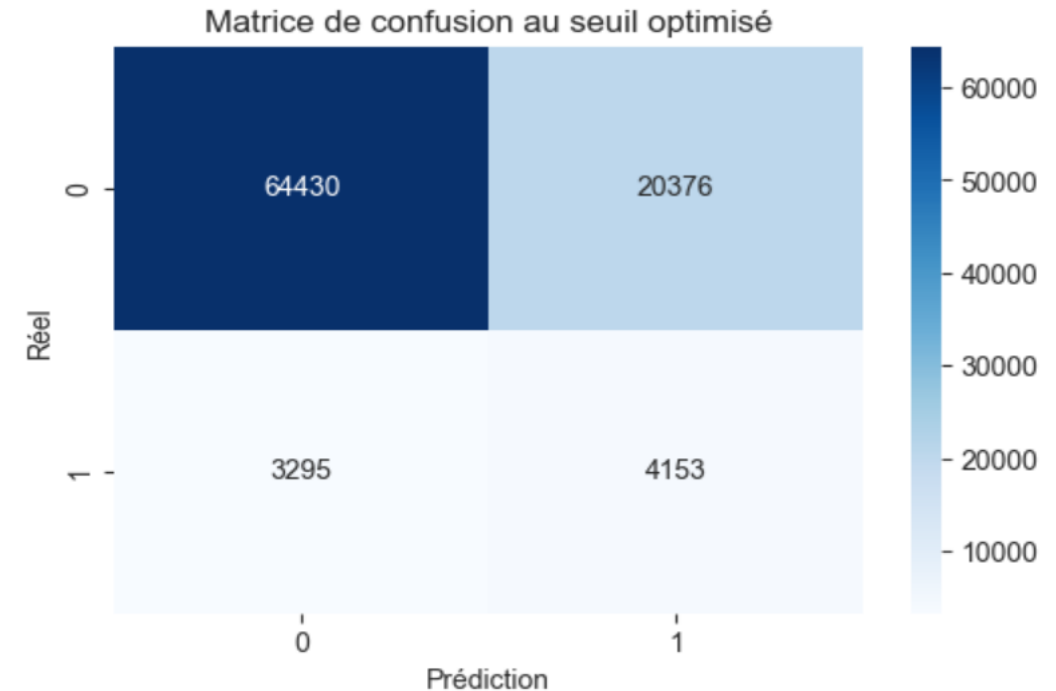
4

Fonction de coût métier

Cette fonction permet de trouver le seuil optimal qui minimise la perte métier

$$F_{cm} = FN * 10 + FP * 1$$

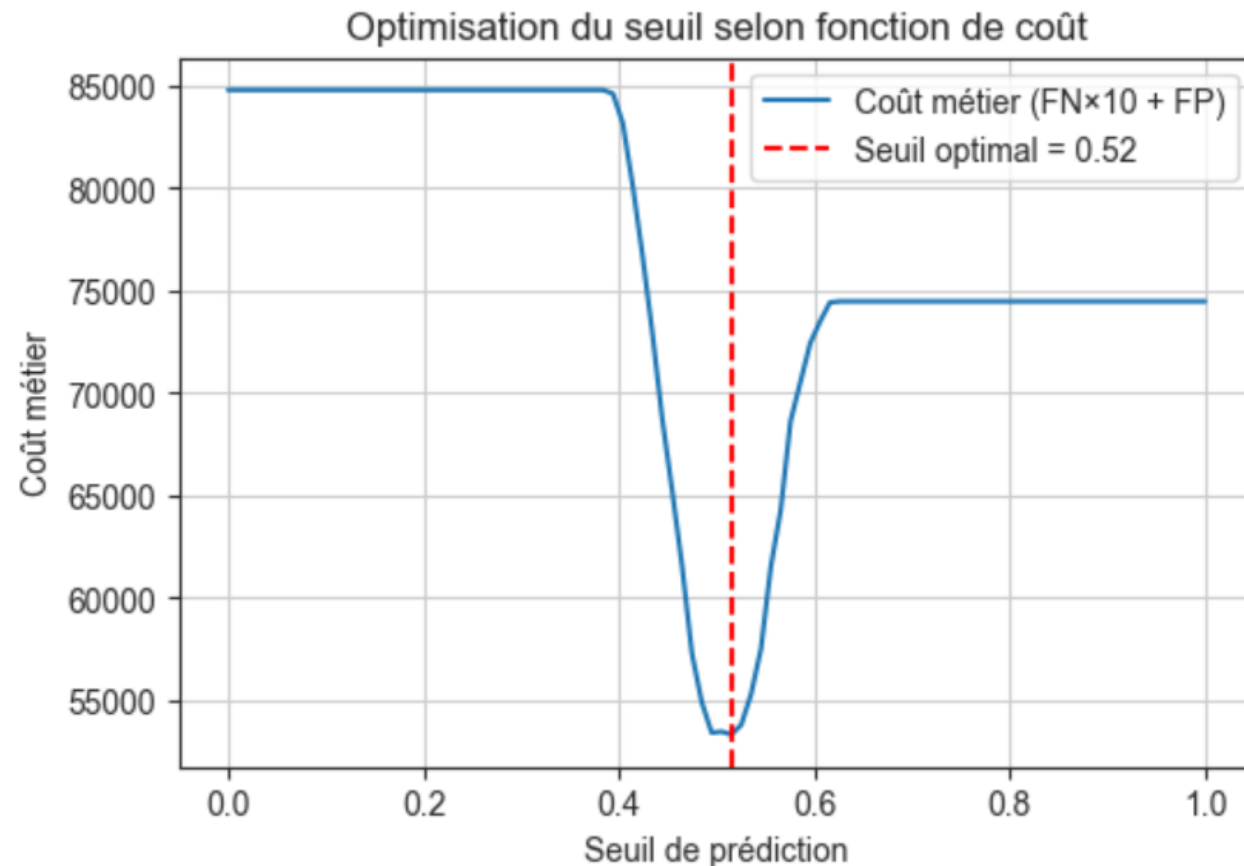
- **FN:** Vrai défaut classé bon client(perte)
- **FP:** Bon client classé comme défaut(manque à gagner)



4

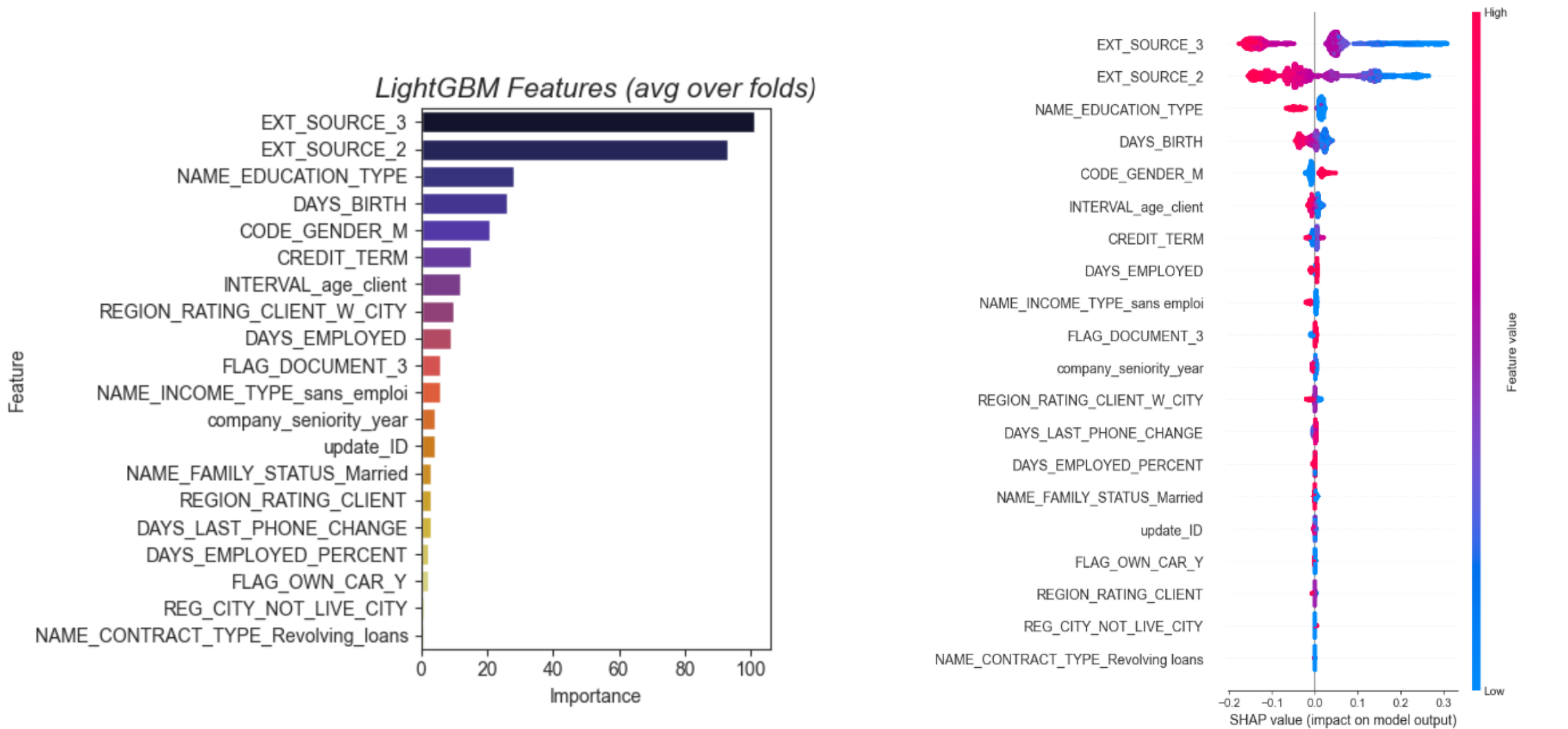
Seuil optimal et coût métier

- Seuil Optimal : **0,515**
- Coût métier minimal : **53326**



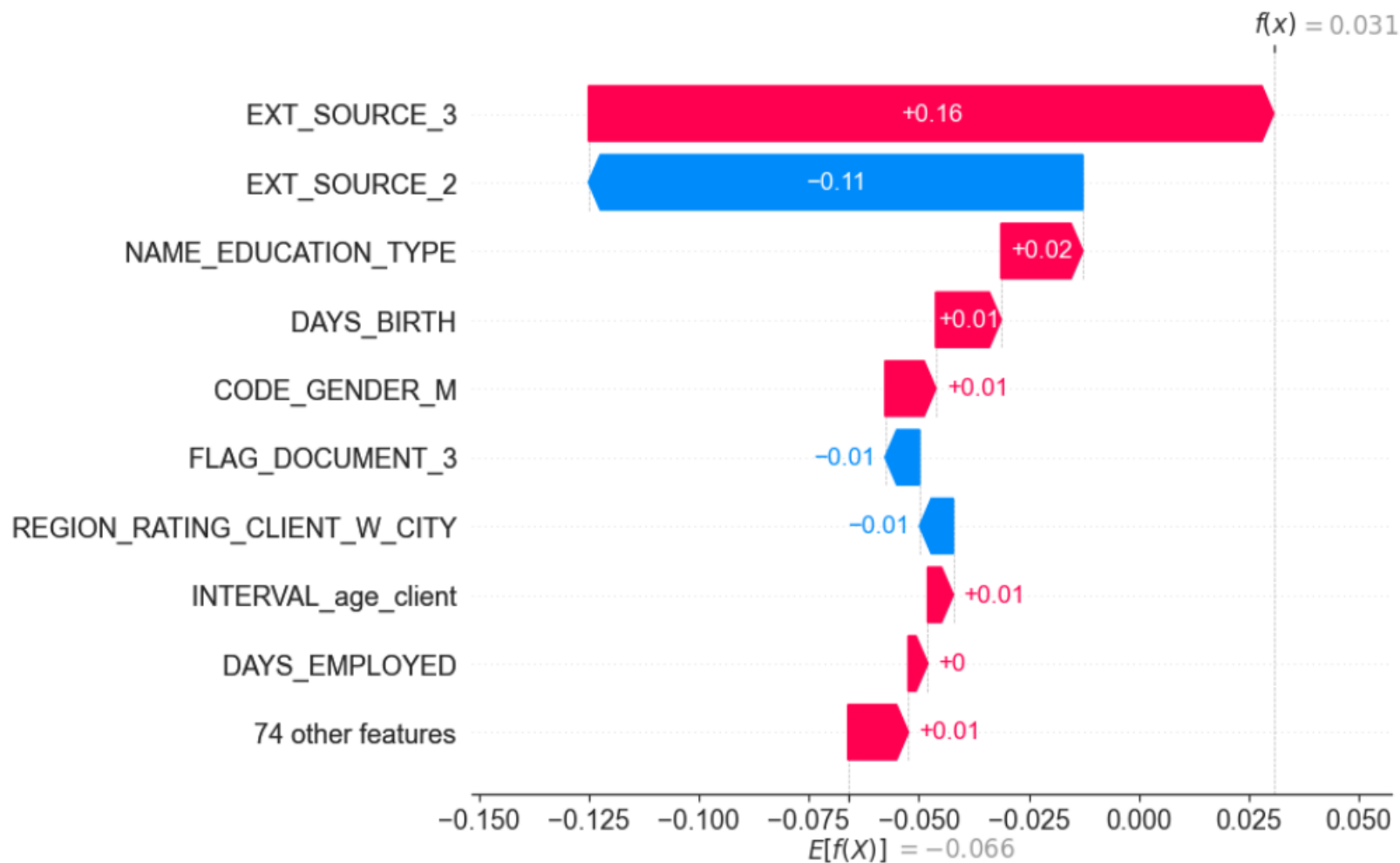
4

Features Importance & Shap globale



5

Shap local : client 15



6

Dashboard

Données du Client

Revenu total

100000.00

Montant du crédit

200000.00

Montant de l'annuité

15000.00

Âge (jours négatifs)

-12000.00

Jours d'emploi

-3000.00

EXT_SOURCE_2

0.50

EXT_SOURCE_3

0.50

Éducation (encodée)

0.0

Souscription au prêt

0.0

Seuil de décision (probabilité max pour accepter le prêt)

0.52

Lancer la prédiction

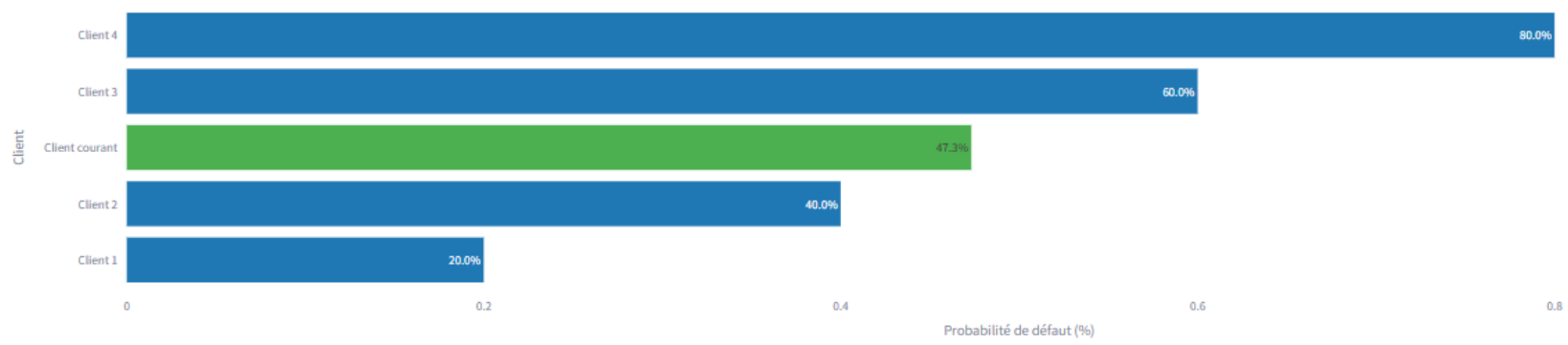
Dashboard de Scoring Crédit

Décision (selon seuil 0.515) : Accordé

Probabilité de non-remboursement
47.32%

Comparaison des probabilités de défaut entre clients

Probabilité de défaut de paiement par client



Mission 2 :

Réalisez une veille technique

Problématique

Réalisez un état de l'art sur une technique récente de modélisation de données texte l'analyser, la tester et la comparer à une approche plus classique que tu as réalisée ?

Contexte du projet

Ce projet consiste à analyser les sentiments exprimés dans des critiques de films issues de la base de données IMDB. L'objectif est de comparer deux approches en traitement automatique du langage naturel (NLP)

Modèle classique: TF-IDF + Naïve

Naïve Bayes : modèle probabiliste basé sur le théorème de Bayes utilisé pour la classification.

Il suppose que chaque mot d'un texte contribue indépendamment au sentiment global du document.

Principe

Pour une critique donnée x , le modèle calcule la probabilité que cette critique appartienne à une classe $y \in \{\text{positive ou négative}\}$.

Vectorisation : TF-IDF

Version améliorée du Bag of words qui pondère les mots selon leur importance dans un Document et dans un document.

Modèle TF-IDF + Naive Bayes

```
-----  
'This movie was a masterpiece.' => positive  
'I was blown away by the acting.' => positive  
'It's a classic that everyone should watch.' => positive  
'The plot was confusing and hard to follow.' => negative  
'The special effects were top-notch.' => negative  
'I couldn't stop laughing throughout the movie.' => negative  
'The soundtrack was incredible.' => positive  
'It's a total waste of time.' => negative  
'I'm still thinking about that ending.' => negative  
'I wouldn't recommend it to anyone.' => positive
```

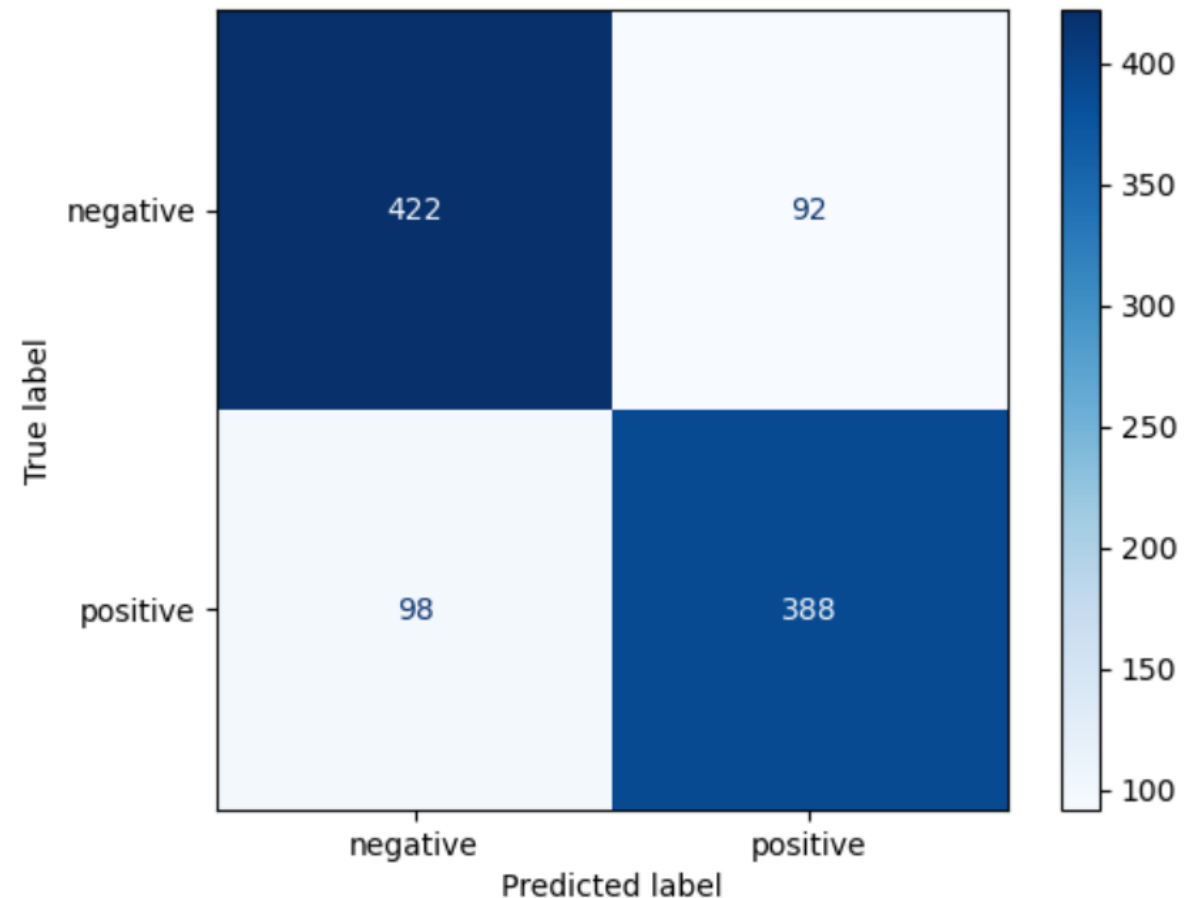
Résultats du modèle classique

Résultats du modèle classique (TF-IDF + Naive Bayes) :

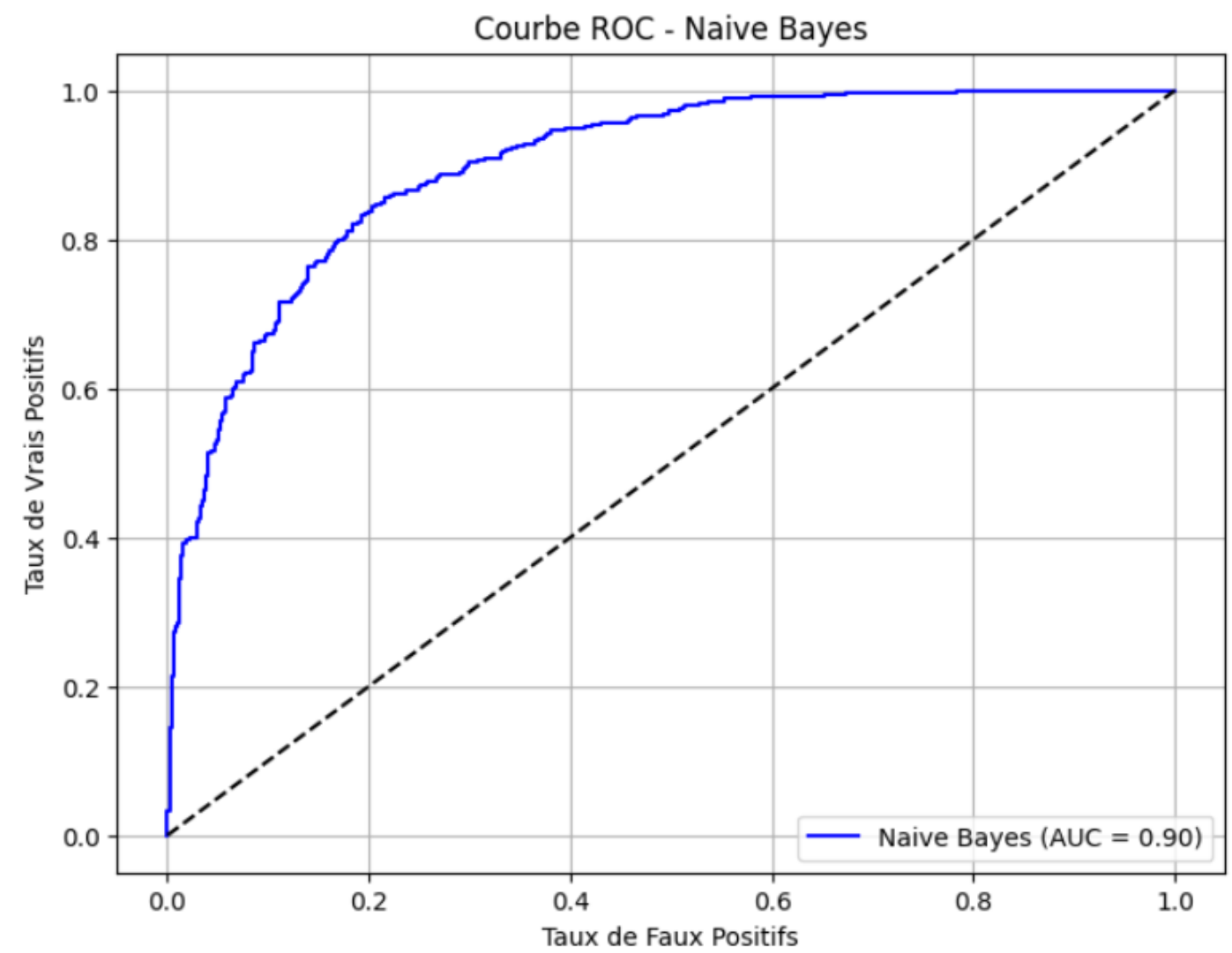
	precision	recall	f1-score	support
negative	0.81	0.82	0.82	514
positive	0.81	0.80	0.80	486
accuracy			0.81	1000
macro avg	0.81	0.81	0.81	1000
weighted avg	0.81	0.81	0.81	1000

Accuracy : 81.00%

Matrice de confusion - TF-IDF + Naive Bayes



Courbe ROC-AUC



Modèle récent : Finetuning complet DistilBert

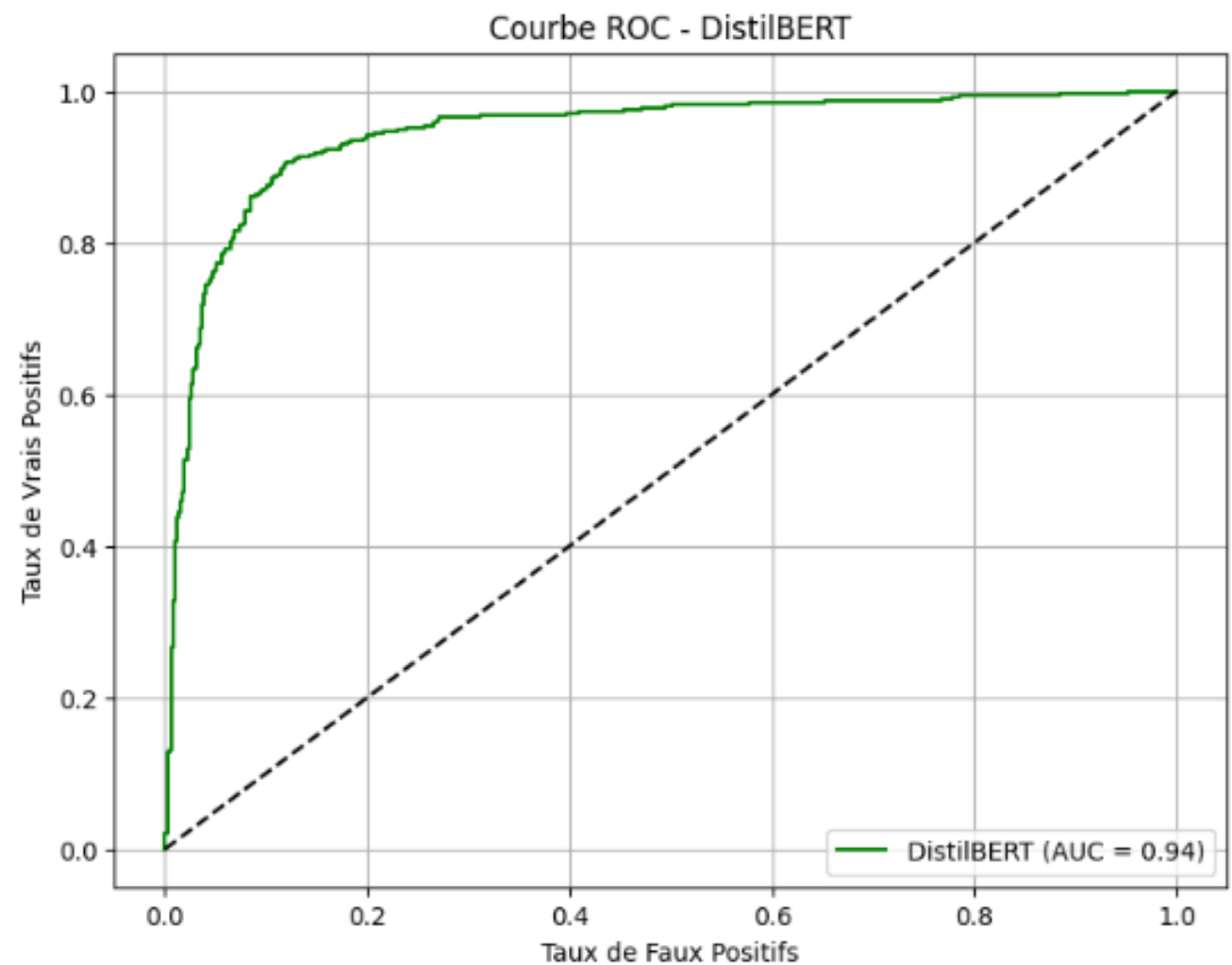
BERT modèle de langage basé sur **l'architecture Transformer**. Il est conçu pour comprendre le contexte d'un mot dans une phrase en lisant à la fois à gauche et à droite du mot.

DistilBERT est une version **plus légère et plus rapide de BERT** Il est obtenu via une méthode appelée **distillation de connaissances**, qui consiste à entraîner un petit modèle (l'“élève”) pour imiter un grand modèle pré-entraîné (le “professeur”, ici BERT).

Source : [arXiv:1910.01108](<https://arxiv.org/abs/1910.01108>)

Résultats du modèle DistilBert

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	No log	0.296848	0.891000	0.890953	0.891122	0.891000
2	No log	0.369752	0.883000	0.882526	0.886713	0.883000
3	No log	0.360713	0.897000	0.897019	0.897557	0.897000



Prédictions du modèle entraîné distilbert:

This movie was a masterpiece. - Positif

I was blown away by the acting. - Négatif

It's a classic that everyone should watch. - Positif

The plot was confusing and hard to follow. - Négatif

The special effects were top-notch. - Positif

I couldn't stop laughing throughout the movie. - Négatif

The soundtrack was incredible. - Positif

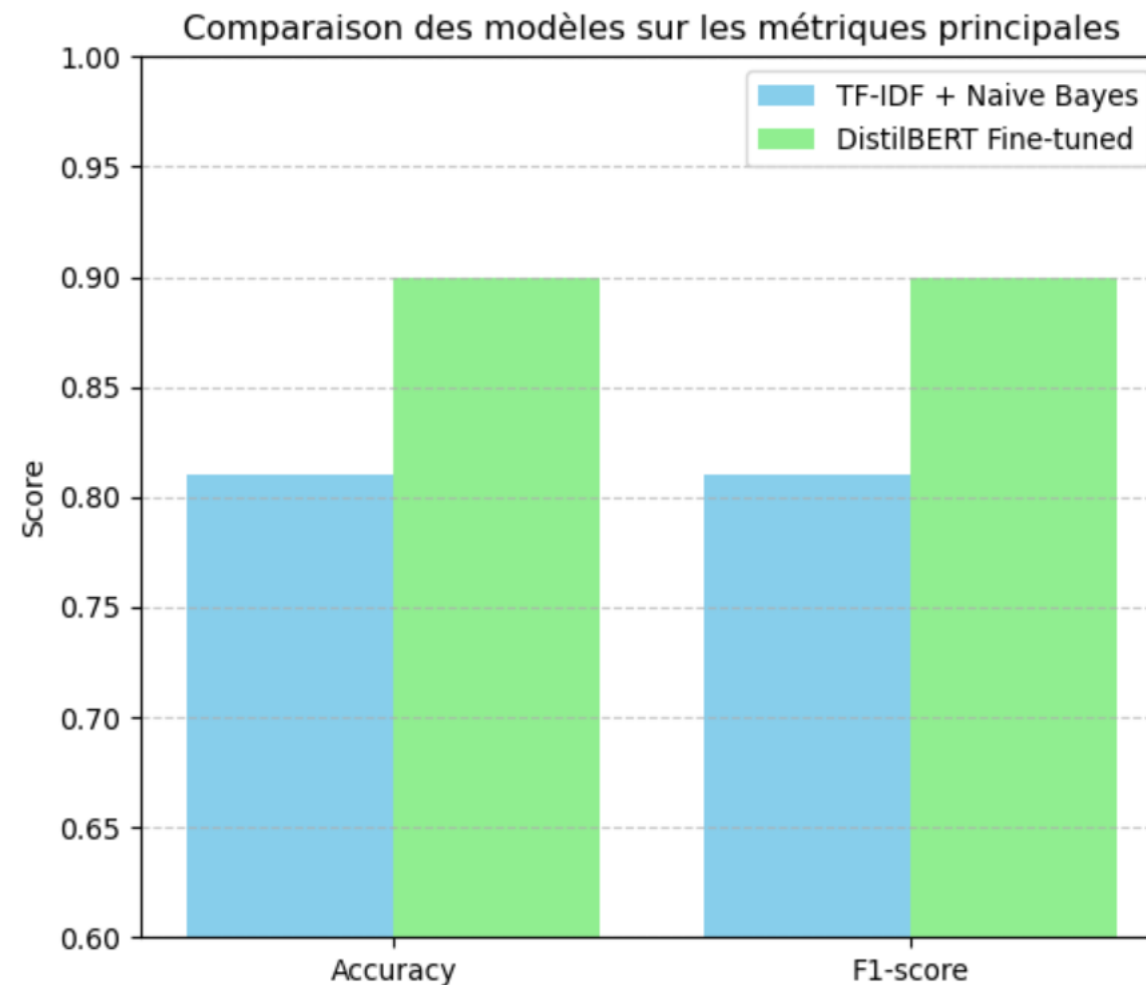
It's a total waste of time. - Négatif

I'm still thinking about that ending. - Négatif

I wouldn't recommend it to anyone. - Négatif

Résultat des comparaisons

Modèles	Accuracy	F1-score
TF-IDF + Naive Bayes	~81%	~0.81
DistilBERT Fine-tuné	90%	0.90



Conclusion

**Merci pour
Votre
Compréhension**